BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

#21
KP
2-19-02
Attachment A

PATENT
025098-1406 (Formerly 238/298)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of:

Baird, *et al.*

Serial No.: 09/374,704

Title: INHIBITION OF MAJOR GROOVE DNA BINDING PROTEINS BY MODIFIED POLYAMIDES

Filing Date: August 12, 1999

Group Art Unit: 1635

Examiner: Janet L. Epps

APPELLANTS' BRIEF IN REPLY TO EXAMINER'S ANSWER

Commissioner for Patents
Washington, D.C. 20231

Sir:

Applicants (herein, "Appellants") submit this Reply Brief pursuant to the Examiner's Answer ("Paper No. 20") mailed on January 2, 2002. No fee is believed due in connection with this submission. If this is incorrect and additional fees are due in this regard, please charge our Deposit Account No. 50-0872 for the appropriate amount.

02/13/2002 GTEFFERA 00000165 09374704
02 FC:120 320.00 OP

209392.1

CERTIFICATE OF MAILING

I hereby certify that this paper (along with any referred to as being attached or enclosed) is being deposited with the United States Postal Service as First Class Mail with sufficient postage in an envelope addressed to the Commissioner for Patents, Washington, D.C. 20231, on the date below.

Jodie M. Price
Name of Person Mailing Paper

January 29, 2002
Date of Deposit

Signature of Person Mailing Paper

Table of Contents

Table of Publications

Grouping of Claims

Appellants respectfully disagree with the Examiner's statement in Paper No. 20, page 2, that Appellants' Brief "does not include a statement that this grouping of claims does not stand or fall together and reasons in support thereof." Such a statement may be found on page 8 of Appellants' Brief.

Argument

Appellants respectfully request that the rejection of claims 1, 3-19, and 25-26 be withdrawn or reversed, and that the instant claims be permitted to proceed to allowance, because the Examiner has failed to establish a *prima facie* case of obviousness. The Examiner's Answer to Appellants' Appeal Brief only serves to emphasize that the Examiner continues to selectively consider the cited publications, ignoring those portions that clearly indicate the flaws in the Examiner's asserted *prima facie* case. Furthermore, the Examiner continues to rely on hindsight in support of the rejection, and has failed to provide any particular findings as to the reason that the skilled artisan, with no knowledge of the claimed invention, would have selected the components that the Examiner seeks to combine.

As discussed in detain in Appellants' Appeal Brief, the Examiner incorrectly contends that it allegedly would have been obvious to select three amino acids – Arg-Pro-Arg – from the much larger Hin recombinase protein (disclosed in the Feng *et al.* publication) for combination with a class of polyamide molecules based on N-methylpyrrole and N-methylimidazole

(disclosed in the Swalley, Parks, and Trauger publications) in order to provide the instantly claimed invention. Appellants described why the Arg-Pro-Arg sequence selected in isolation by the Examiner is neither chemically nor functionally similar to such pyrrole- and imidazole-based polyamides disclosed in the Swalley, Parks, and Trauger publications. Moreover, Appellants also described why the skilled artisan would readily acknowledge that the Arg-Pro-Arg sequence in isolation does not possess any DNA binding characteristics. Rather, it is only in the context of a much larger and well known DNA binding motif that the "Arg-Pro-Arg" sequence selected by the Examiner has any DNA binding characteristics whatsoever.

Therefore, Appellants pointed out that one of ordinary skill in the art at the time the instant application was filed would have lacked any motivation to select the Arg-Pro-Arg sequence from a much larger protein structure, in order to graft that sequence onto a pyrrole- and imidazole-based polyamide to provide the instantly claimed invention; and that one of ordinary skill in the art would have also lacked a reasonable expectation that grafting an Arg-Pro-Arg sequence onto such polyamides would have any effect whatsoever on the ability of such polyamides to bind DNA.

The Examiner continues to fail to consider the teachings of cited publications in their entirety with the knowledge available to the skilled artisan at the time the instant application was filed

In attempting to rebut the arguments made by Appellants in their Appeal Brief, the Examiner continues to selectively consider only certain portions of the cited publications, while ignoring those portions that clearly demonstrate the fallacy of the Examiner's rebuttal. For example, Appellants discussed in detail why the skilled artisan would clearly understand from the cited Feng *et al.* publication that it is only in the context of the larger Hin recombinase

protein that the Arg-Pro-Arg sequence exhibit any DNA binding properties whatsoever, and that the skilled artisan would not have excised these three residues from the much larger Hin recombinase protein for grafting onto a pyrrole- or imidazole-based polyamide. The Examiner's statement in reply that "the Feng et al. reference clearly states that the 'Arg-Pro-Arg' sequence maintains its ability to interact with the minor groove of DNA in a specific manner when positioned in an entirely different protein sequence, namely the *Engrailed* protein of *Drosophila*" (Paper No. 20, page 6) fails to consider that the *Engrailed* protein and the Hin recombinase protein are not "entirely different," as the Examiner contends. Instead, each of these two proteins rely on the very same well known structurally conserved helix-turn-helix DNA binding motif to provide specific binding to DNA.

As described in the Feng *et al.* publication on page 348, right column, specific binding of DNA by Hin recombinase "requires both major groove interactions involving a helix-turn-helix (HTH) α-helix motif and minor groove interactions involving the sequence Gly^{139}-Arg^{140}- Pro^{141}-Arg^{142}" (emphasis added). The helix-turn-helix motif is known by those of skill in the art to provide sequence specific DNA binding. *See, e.g.*, The Dictionary of Cell Biology, 2nd Ed. (Lackie and Dow, 1995) (helix-turn-helix: "A motif associated with transcription factors, allowing them to bind to and recognize specific DNA sequences"). As shown in Figure 3 of the Feng *et al.* publication, the helix-turn-helix motif serves to present the Gly^{139}-Arg^{140}- Pro^{141}-Arg^{142} sequence in precisely the proper orientation to interact with the DNA molecule.

Similarly, *Engrailed* is a homeodomain protein that contains the very same helix-turn-helix motif as Hin recombinase. *See*, Feng *et al.*, Figure 10 (showing engrailed homeodomain); *see also*, The Penguin Dictionary of Biology, 10th Ed. (2000) (homeobox: "Conserved DNA

sequence encoding DNA binding regions of many proteins, [including] homeodomains... [conferring] the helix-turn-helix motif upon a protein, giving it its DNA binding properties"). As shown in Figure 10 of the Feng *et al.* publication, the helix-turn-helix motif of Hin recombinase and *Engrailed* interact with DNA in a highly conserved fashion.

Thus, based on the well understood structure and function of the helix-turn-helix motif in proteins such as Hin recombinase and *Engrailed*, the skilled artisan would understand that it is only in the context of the strikingly similar 3-dimensional structure of these proteins that the Arg-Pro-Arg sequence selected by the Examiner exhibits any DNA binding characteristics whatsoever. The Examiner's contention that "the prior art has provided two examples wherein the sequence 'Arg-Pro-Arg' is instrumental in providing minor groove specific binding to DNA" (Paper No. 20, page 8) is absurdly reductionist, ignoring the context in which the skilled artisan would view the Feng *et al.* publication. Indeed, taken to its logical conclusion, the Examiner's argument would entitle the Examiner to assert that even a single amino acid in a protein is "instrumental" to the function of that protein.

The Examiner has steadfastly ignored the fact that, as recognized in the Feng *et al.* publication, DNA binding requires a large number of contacts, both within the protein itself (*e.g.*, formation of the helix-turn-helix motif), and between the protein and DNA (*e.g.*, in both in both the major and minor DNA groove), and is not conferred by a single 3-residue peptide. Nothing in the purported *prima facie* case would indicate to one of skill in the art that, in the absence of this helix-turn-helix motif, the "Arg-Pro-Arg" sequence would exhibit any DNA binding properties whatsoever, and nothing of record indicates that this helix-turn-helix motif would exist in the pyrrole- and imidazole-based polyamides of the Swalley, Parks, and Trauger publication.

Therefore, the skilled artisan would not be motivated to combine an Arg-Pro-Arg sequence to pyrrole- and imidazole-based polyamides in order to provide the instantly claimed invention, or have a reasonable expectation that grafting the Arg-Pro-Arg sequence onto a pyrrole- or imidazole-based polyamide would have any effect whatsoever on the ability of such polyamides to bind DNA.

The Examiner also fails to provide any reasons why the skilled artisan, with no knowledge of the claimed invention, would have selected the components selected by the Examiner

Furthermore, in the face of a clear statement in the Feng *et al.* publication that specific binding of DNA by Hin recombinase "requires both major groove interactions involving a helix-turn-helix (HTH) α-helix motif and minor groove interactions involving the sequence Gly^{139}-Arg^{140}- Pro^{141}-Arg^{142}," the Examiner has provided no evidence to support the contention that the skilled artisan would allegedly be motivated to select the particular three amino acid segment chosen by the Examiner in isolation from the much larger 52 amino acid Hin recombinase DNA binding domain. *See, In re Dance*, 48 USPQ2d 1635, 1637 (Fed. Cir. 1998) (there must be some motivation, suggestion, or teaching of the desirability of making the specific combination that was made by applicant).

For example, while the Feng *et al.* publication provides no information about the Arg-Pro-Arg sequence selected in isolation by the Examiner, the publication does state that the Gly residue in the sequence Gly_{139}-Arg_{140} is essential for sequence specific DNA binding by Hin recombinase. *See*, Feng *et al.*, page 348, column 3. Why then would the skilled artisan be motivated to select the Arg-Pro-Arg sequence without the Gly residue? The Examiner is silent on this point.

Similarly, the Feng *et al.* publication also states that α-helix 3, which is located at the carboxyl-terminal end of the DNA binding domain, is the DNA recognition helix for the Hin recombinase protein. Feng *et al.*, page 350, left column; Fig. 1, p. 348. Why would the skilled artisan be motivated to select the Arg-Pro-Arg sequence and not the sequence in α-helix 3? Again, the Examiner is silent.

Additionally, the Feng *et al.* publication states that the Gly_{139}-Arg_{140}-Pro_{141}-Arg_{142} sequence adopts an extended conformation to present the 4 amino acid sequence in a proper manner to bind DNA, and that Ile_{144} is critical in maintaining this orientation. Feng *et al.*, page 351, columns 2-3. Why would the skilled artisan be motivated to select the Arg-Pro-Arg sequence and not a sequence that includes this Ile residue? Again, silence by the Examiner.

Moreover, as discussed above, the skilled artisan understands that the helix-turn-helix motif itself can confer sequence-specific DNA binding properties on a protein. Why would the skilled artisan be motivated to select the Arg-Pro-Arg sequence and not all or a portion of the helix-turn-helix motif? Again, silence.

Finally, the skilled artisan also understands that there are numerous DNA binding proteins that use a motif other than the helix-turn-helix motif, such as the helix-loop-helix motif, the leucine zipper motif, and the zinc finger motif. *See, e.g.*, Dictionary of Cell Biology definitions of each of these motifs. Why would the skilled artisan be motivated to zero in on only the helix-turn-helix proteins, such as Hin recombinase, at all? The Examiner provides no explanation.

Viewed in this light, it is apparent that the only means by which the Examiner has selected the Arg-Pro-Arg sequence is by hindsight, "decomposing an invention into its

constituent elements, finding each element in the prior art, and then claiming that it is easy to reassemble these elements into the invention." *In re Mahurkar*, 28 USPQ2d 1801, 1817 (N.D. Ill. 1993). The Court of Appeals for the Federal Circuit ("CAFC") has repeatedly counseled that "the best defense against the subtle but powerful attraction of a hindsight-based obviousness analysis is rigorous application of the requirement for a showing of the teaching or motivation to combine prior art references." *In re Dembiczak*, 50 USPQ2d 1614, 1617 (Fed. Cir. 1999).

The Examiner's statement that "any judgement on obviousness is in a sense necessarily a reconstruction based upon hindsight reasoning" (Paper No. 20, page 9) exhibits a complete lack of understanding by the Examiner that "particular findings must be made as to the reason the skilled artisan, with no knowledge of the claimed invention, would have selected these components for combination in the manner claimed" in order to support an obviousness rejection. *In re Kotzab*, 55 USPQ2d 1313, 1317 (Fed. Cir. 2000) (emphasis added). Because the Examiner has failed to provide such particular findings, no *prima facie* case of obviousness has been established.

In contradiction to the Examiner's contention, the Arg-Pro-Arg sequence excised by the Examiner from Hin recombinase is not functionally or structurally similar to the polyamides disclosed in the Swalley, Parks, and Trauger publications

Furthermore, the Examiner's axiomatic statement that "it is *prima facie* obvious to combine two compositions each of which is taught by the prior art to be useful for the same purpose" (Paper No. 20, page 8) can provide no solace to the Examiner. As discussed above, the Examiner's contention that "the prior art has provided two examples wherein the sequence 'Arg-Pro-Arg' is instrumental in providing minor groove specific binding to DNA" is reductionist in the extreme, and would logically lead to the absurd conclusion that even a single amino acid in

the Hin recombinase structure is "instrumental" to DNA binding (and apparently that the skilled artisan would be motivated to select it over any other residue or combination of residues in the protein).

In contrast to the Arg-Pro-Arg sequence selected by the Examiner, the pyrrole- and imidazole-based polyamides disclosed in the Swalley, Parks, and Trauger publications were actually demonstrated to specifically bind to DNA in and of themselves. Thus, in contradiction to the Examiner's contention, nothing in the purported *prima facie* case indicates that any sequence (other than perhaps the entire Hin recombinase DNA binding domain) is "useful for the same purpose" as the polyamides disclosed in the Swalley, Parks, and Trauger publications. Similarly, nothing in the purported *prima facie* case indicates that "the arginine and proline residues are disclosed in Feng et al. [are] functionally similar to the polyamide structures of Swalley, Parks, and Trauger." Paper No. 20, page 7.

The Examiner has attempted to bolster the fatally flawed *prima facie* case with two newly introduced arguments. First, the Examiner contends that "in both the Hin recombinase and the *Engrailed* protein, the Arg-Pro-Arg sequence is positioned in the amino terminus, and it is in this environment that minor groove specific binding of DNA occurs." Unfortunately for the Examiner, however, the Arg-Pro-Arg sequence is not positioned at the amino terminus of either Hin recombinase or *Engrailed*. In *Engrailed*, the residues are in positions 3, 4, and 5; while in Hin recombinase, the residues are in positions 140, 141, and 142, where position 1 represents the amino terminus. The Arg-Pro-Arg residues do, however, appear to lie at or near the amino group end of the helix-loop-helix motif (*see, e.g.*, Feng *et al.*, Figure 3), again emphasizing that it is the

structure of the DNA binding domain in its entirety, and not some isolated property of the Arg-Pro-Arg sequence itself, that provides DNA binding properties to Hin recombinase.

Second, referring to only Arg^{140}, the Examiner also contends that "Feng et al. teaches that the ε-amine of the Arg side chain is capable of interacting with the N-3 group of Adenine, and the main chain Arg permits a second hydrogen bond with the O-2 of Thymine," concluding that the structure of Arg "permits a stable interaction with A,T base pairs [in similar fashion to the] side-by-side pairing in the polyamides of Parks *et al.*, Swalley *et al.*, and Trauger *et al.*" Paper No. 20, page 7.

Once again, however, the Examiner's arguments only serve to emphasize the hindsight reconstruction being employed by the Examiner to justify the flawed obviousness rejection. The Examiner provides no insight as to why the skilled artisan would focus on this single amino acid, while ignoring the fact that Gly^{139} also makes contacts with Adenine, and that substitution of Adenine with Guanine would disrupt those contacts. Feng *et al.*, page 351, right column. Similarly, one is left without an explanation as to why the skilled artisan would ignore the final six residues Ile^{185}-Lys^{186}-Lys^{187}-Arg^{188}-Met^{189}-Asn^{190}, when the Feng *et al.* publication indicates that the binding of these residues to DNA "recalls AT-specific binding of minor groove drugs such as netropsin and distamycin." Feng *et al.*, page 352, left column. Indeed, the Feng *et al.* publication is replete with description of various Hin recombinase residues that bind to DNA and the contacts made by those residues. Thus, the Examiner again selectively considers only specific portions of the Feng *et al.* publication, ignoring numerous sections that clearly demonstrate the flaws in the asserted *prima facie* case.

The Examiner has failed to considered the problem to be solved by the instant invention, rendering the obvious analysis fatally flawed

The Court of Appeals for the Federal Circuit has counseled on the importance of considering the problem to be solved in any obviousness determination. *See, e.g., Sibia Neurosciences, Inc. v. Cadus Pharmaceutical Corp.*, 55 USPQ2d 1927, 1933 (Fed. Cir. 2000). Because the Examiner has failed to properly consider the problem to be solved by the instant invention, the obviousness analysis performed by the Examiner is fatally flawed and cannot stand.

The Examiner's asserted *prima facie* case is based on an alleged suggestion in the Feng *et al.* publication that "the Arg-Pro-Arg sequence would maintain specific minor groove binding especially when positioned at a terminus of the molecule." Paper No. 20, page 6. But, as pointed out in the Appeal Brief, the "positive patch" of the instant claims is not intended to serve a minor groove binding function. Rather, the positively charged group is intended to disrupt interactions between proteins and the phosphate backbone or major groove of the DNA molecule. *See, e.g.*, specification, page 14, lines 20-29. There is nothing of record in the asserted *prima facie* case to indicate that the skilled artisan would be motivated to select an Arg-Pro-Arg sequence, as disclosed in the Feng *et al.* publication, in order to disrupt interactions between proteins and the phosphate backbone or major groove of the DNA molecule.

The Examiner's axiomatic statement in reply that "[a]lthough the claims are interpreted in light of the specification, limitations from the specification are not read into the claims" (Paper No. 20, page 9) completely ignores the extremely relevant question of whether the skilled artisan, seeking to design molecules that disrupt interactions between proteins and the phosphate

backbone or major groove of the DNA molecule, would have been led to consider the combination of features proposed by the Examiner. Further, the Examiner's reply indicates a failure to understand the importance of considering the problem to be solved in performing an obviousness determination. Because of this failure, no properly supported *prima facie* case of obviousness has been demonstrated.

Conclusion

In view of the foregoing, Appellants respectfully submit that the Examiner's asserted *prima facie* case of obviousness is based upon a selective, piecemeal reading of the cited publications and a failure to understand and perform the necessary analysis or to provide sufficient findings on which a proper obviousness rejection may be based. Instead, it becomes clear that the Examiner has relied solely on hindsight in selecting certain components, and ignoring others, for combination. Therefore, the Examiner has failed to carry the burden of establishing a *prima facie* case of obviousness. Appellants respectfully submit the instant claims are in condition for allowance, and respectfully request that the rejections be withdrawn or reversed, and that the rejected claims, together with the claims previously indicated as allowable, be allowed to issue.

Respectfully submitted,

FOLEY & LARDNER

Dated: January 29, 2002

By: _____
Michael A. Whittaker
Registration No. 46,230

P.O. Box 80278
San Diego, CA 92138-0278
Telephone: (858) 847-6700

J. Am. Chem. Soc. 1996, 118, 8198−8206

Recognition of a 5′-(A,T)GGG(A,T)$_2$-3′ Sequence in the Minor Groove of DNA by an Eight-Ring Hairpin Polyamide

Susanne E. Swalley, Eldon E. Baird, and Peter B. Dervan*

Contribution from the Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125

Received April 8, 1996[⊗]

Abstract: The use of pyrrole−imidazole polyamides for the recognition of core 5′-GGG-3′ sequences in the minor groove of double stranded DNA is described. Two hairpin pyrrole−imidazole polyamides, ImImIm-γ-PyPyPy-β-Dp and ImImImPy-γ-PyPyPyPy-β-Dp (Im = *N*-methylimidazole-2-carboxamide, Py = *N*-methylpyrrole-2-carboxamide, β = β-alanine, γ = γ-aminobutyric acid, and Dp = ((dimethylamino)propyl)amide), as well as the corresponding EDTA affinity cleaving derivatives, were synthesized and their DNA binding properties analyzed. Quantitative DNase I footprint titrations demonstrate that ImImIm-γ-PyPyPy-β-Dp binds the formal match sequence 5′-AGGGA-3′ with an equilibrium association constant of $K_a = 5 \times 10^6$ M^{-1} (10 mM Tris·HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0 and 22 °C). ImImImPy-γ-PyPyPyPy-β-Dp binds the same site, 5′-AGGGAA-3′, approximately two orders of magnitude more tightly than the six ring polyamide, with an equilibrium association constant of $K_a = 4 \times 10^8$ M^{-1}. The eight-ring hairpin polyamide demonstrates greater specifity for single base pair mismatches than does the six ring hairpin. Polyamides with an EDTA·Fe(II) moiety at the carboxy terminus confirm that each hairpin binds in a single orientation. The high affinity recognition of a 5′-GGG-3′ core sequence by an eight ring polyamide containing three contiguous imidazole amino acids demonstrates the versatility of pyrrole−imidazole polyamides and broadens the sequence repertoire for DNA recognition.

Introduction

Pyrrole−imidazole polyamide−DNA complexes afford a general method for the design of non-natural molecules for sequence-specific recognition in the minor groove of DNA.[1-5] Polyamides containing *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) carboxamides bind to the minor groove as side-by-side antiparallel dimers and are capable of specific recognition of sequences containing G,C base pairs, where the N3 of each imidazole forms a hydrogen bond with a single guanine exocyclic amino group.[1] The side-by-side pairing of an imidazole ring from one polyamide with a pyrrole ring from the second polyamide recognizes a G·C base pair, while a pyrrole−imidazole combination recognizes a C·G base pair.[1] Finally, a pyrrole−pyrrole pair recognizes either an A·T or T·A base pair.[1,2] By employing the 2:1 model, specific recognition of the sequences 5′-(A,T)G(A,T)C(A,T)-3′, 5′-(A,T)G(A,T)$_3$-3′, 5′-(A,T)$_2$G(A,T)$_2$-3′, and 5′-(A,T)GCGC(A,T)-3′ has been achieved.[1-5]

Covalent head-to-tail linkage of two polyamides by γ-aminobutyric acid (γ) to form a "hairpin" polyamide results in both increased affinity and specificity, as compared to the unlinked polyamides.[6] For instance, the 1:1 hairpin motif has been used to recognize 5′-(A,T)G(A,T)$_3$-3′ by ImPyPy-γ-PyPyPy-Dp with approximately 300-fold greater affinity than the unlinked polyamides, ImPyPy and PyPyPy.[6] A C-terminal β-alanine residue increases both affinity and specificity and facilitates solid phase synthesis, as recently demonstrated with ImPyPy-γ-PyPyPy-β-Dp.[7,8] Furthermore, a sequence containing two contiguous G·C base pairs, 5′-(A,T)GG(A,T)$_2$-3′, has been recognized by ImImPy-γ-PyPyPy-β-Dp.[9]

To further expand the sequence repertoire available with the hairpin motif, two polyamides containing three contiguous imidazole rings, ImImIm-γ-PyPyPy-β-Dp (**1**) and ImImImPy-γ-PyPyPyPy-β-Dp (**2**), and the corresponding affinity cleaving analogs, ImImIm-γ-PyPyPy-β-Dp-EDTA (**1-E**) and ImImImPy-γ-PyPyPyPy-β-Dp-EDTA (**2-E**), were synthesized using solid phase synthetic protocols (Figures 1 and 2).[7] Specific hydrogen bonds are expected to form between each imidazole N3 and one of the three individual guanine 2-amino groups on the floor of the minor groove (Figure 1). The eight-ring hairpin polyamide, with a pyrrole between the C-terminal imidazole and the γ-linker, was synthesized to examine whether the positioning of the final imidazole immediately adjacent to the turn would adversely affect binding affinity or specificity. We report here the affinities and relative specificities of these tris-imidazole polyamides as determined by three separate techniques: MPE·Fe(II) footprinting,[10] DNase I footprinting,[11] and affinity cleaving.[12] Information about binding site size is gained from MPE·Fe(II) footprinting, while quantitative DNase I footprint titrations

[⊗] Abstract published in *Advance ACS Abstracts,* August 15, 1996.

(1) (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* 1992, *114*, 8783. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* 1992, *89*, 7586. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* 1993, *32*, 11385.

(2) (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* 1989, *86*, 5723. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* 1990, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingham, M. *Struct. Biol. Nature* 1994, *1*, 169.

(3) (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* 1993, *115*, 2572. (b) Geierstanger, B. H.; Jacobsen, J.-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* 1994, *33*, 3055.

(4) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* 1993, *115*, 4474.

(5) (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* 1994, *266*, 646. (b) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* 1995, *117*, 3325.

(6) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* 1994, *116*, 7983.

(7) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* 1996, *118*, 6141.

(8) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* 1996, *118*, 6147.

(9) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* 1996, *118*, 6153.

(10) (a) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* 1983, *22*, 2373. (b) Van Dyke, M. W.; Dervan, P. B. *Nucleic Acids Res.* 1983, *11*, 5555.

S0002-7863(96)01159-6 CCC: $12.00

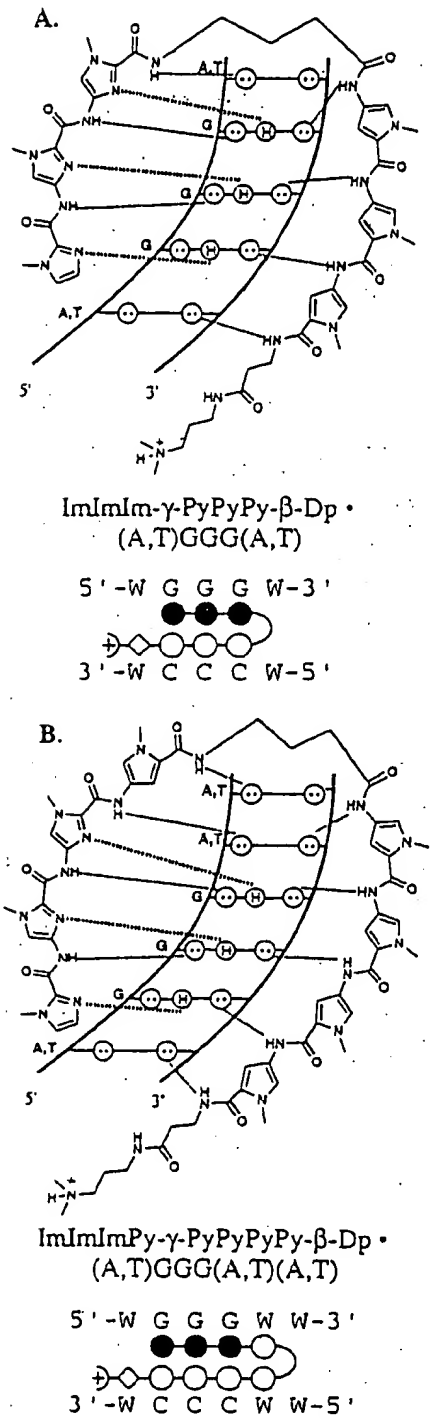


Figure 1. Binding model for the complexes formed between the DNA and either the six-ring hairpin polyamide ImImIm-γ-PyPyPy-β-Dp (A) or the eight-ring hairpin polyamide ImImImPy-γ-PyPyPyPy-β-Dp (B). Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. Ball and stick models are also shown. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent the β-alanine residue. W represents either an A or T base.

allow the determination of equilibrium association constants (K_a) for the polyamides to a variety of match and single base pair mismatch sequences. Affinity cleavage studies confirm the binding orientation and stoichiometry of the 1:1 hairpin:DNA complex.

Results

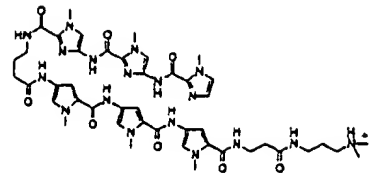
Synthesis of Polyamides. The polyamides ImImIm-γ-PyPyPy-β-Dp (1) and ImImImPy-γ-PyPyPyPy-β-Dp (2) were synthesized in a stepwise manner from Boc-β-alanine-Pam resin using recently described Boc-chemistry protocols in 14 and 18 steps, respectively.[7] The polyamides were then cleaved by a single step aminolysis reaction with ((dimethylamino)propyl)-amine and subsequently purified by HPLC chromatography. The syntheses of the polyamides, ImImImPy-γ-PyPyPyPy-β-Dp (2), ImImImPy-γ-PyPyPyPy-β-Dp-NH_2 (2-NH_2), and ImImImPy-γ-PyPyPyPy-β-Dp-EDTA (2-E), are outlined (Figure 3). For the synthesis of the EDTA analog, a sample of resin is treated with 3,3'-diamino-*N*-methyldipropylamine (55 °C) and then purified by preparatory HPLC to provide 2-NH_2. The polyamide ImImImPy-γ-PyPyPyPy-β-Dp-NH_2 (2-NH_2) provides a free aliphatic primary amine group suitable for modification. This polyamide-amine is then treated with an excess of the dianhydride of EDTA (DMSO/NMP, DIEA, 55 °C) and the remaining anhydride hydrolyzed (0.1 M NaOH, 55 °C). The EDTA modified polyamide is then isolated by preparatory HPLC.

Identification of Binding Site Size and Orientation by MPE·Fe(II) Footprinting and Affinity Cleaving. MPE·Fe(II) footprinting on the 3'- and 5'-^{32}P end-labeled 274 base pair *Eco*RI/*Pvu*II restriction fragment from plasmid pSES1 (25 mM Tris—acetate, 10 mM NaCl, 100 μM calf thymus DNA, pH 7.0 and 22 °C) reveals that the polyamides, each at 10 μM concentration, are binding to the four designed sites, 5'-AGGGA(A)-3', 5'-AGGCA(A)-3', 5'-TGGGT(C)-3', and 5'-TGGGC(T)-3' (Figures 4 and 5). No footprinting is seen with either compound at the single base pair mismatch site, 5'-AGGAA(A)-3' (Figure 4, quantitated data not shown). The footprinting patterns for the six-ring polyamide are consistent with five base pair binding sites, while the patterns for the eight-ring polyamide are indicative of six base pair binding sites. Affinity cleavage experiments on the same restriction fragment (25 mM Tris—acetate, 200 mM NaCl, 50 μg/mL glycogen, pH 7.0 and 22 °C) by the six-ring and eight-ring EDTA·Fe(II) analogs reveal cleavage patterns that are 3'-shifted and appear on only one side of each binding site (Figures 4 and 5). ImImIm-γ-PyPyPy-β-Dp-EDTA·Fe(II), at 1 μM, and ImImImPy-γ-PyPyPyPy-β-Dp-EDTA·Fe(II), at 100 nM, show cleavage patterns that demonstrate recognition of the same binding sites identified by MPE·Fe(II) footprinting. No carrier DNA was used in these experiments, and thus the relative cleavage intensities indicate that the six-ring polyamide binds most strongly to the two match sites 5'-AGGGA-3' and 5'-TGGGT-3'; followed by the end mismatch 5'-TGGGC-3'. The core mismatch 5'-AGGCA-3', with little appreciable cleavage at 1 μM concentration, is bound more weakly. Similarly, the eight-ring polyamide binds most strongly at 100 nM to 5'-AGGGAA-3', much less strongly to 5'-TGGGTC-3', and not significantly to 5'-TGGGCT-3' and 5'-AGGCAA-3'.

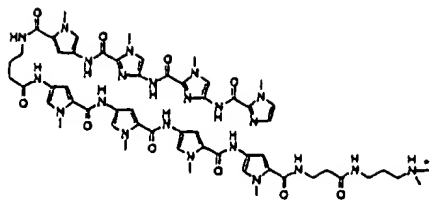
Determination of Binding Affinities by Quantitative DNase I Footprinting. Quantitative DNase I footprint titration experiments (10 mM Tris—HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0 and 22 °C) were performed to determine

(11) (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* 1986, *130*, 132. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. U.S.A.* 1986, *83*, 8462. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* 1986, *25*, 7344.

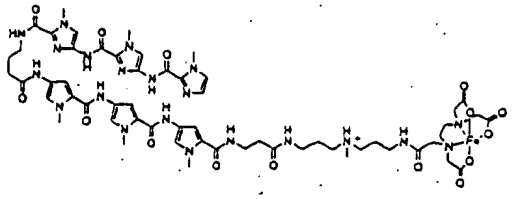
(12) (a) Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* 1982, *104*, 6861. (b) Schultz, P. G.; Dervan, P. B. *J. Biomol. Struct. Dyn.* 1984, *1*, 1133. (c) Taylor, J. S.; Schultz, P. B.; Dervan, P. B. *Tetrahedron* 1984, *40*, 457.



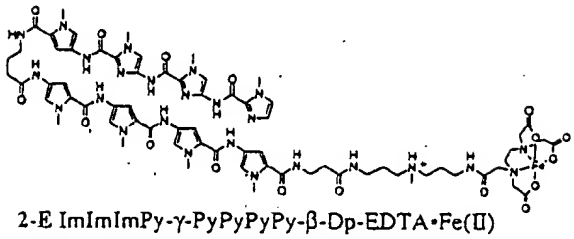
1 ImImIm-γ-PyPyPy-β-Dp



2 ImImImPy-γ-PyPyPyPy-β-Dp



1-E ImImIm-γ-PyPyPy-β-Dp-EDTA•Fe(II)



2-E ImImImPy-γ-PyPyPyPy-β-Dp-EDTA•Fe(II)

Figure 2. Structures of the tris-imidazole polyamides and their EDTA derivatives synthesized using solid-phase methodology.[7]

the equilibrium association constants of the polyamides for the four bound sites (Table 1).[11] ImImIm-γ-PyPyPy-β-Dp binds the two match sites 5′-AGGGA-3′ and 5′-TGGGT-3′ with equilibrium association constants of $K_a = 4.6 \times 10^6$ and 7.6×10^6 M^{-1}, respectively (Figures 6–8). The sequence 5′-TGGGC-3′, which has a mismatch in the non-core final position (an "end" mismatch), is bound with 6-fold lower affinity than the best match, while the core mismatch sequence 5′-AGGCA-3′ is bound with 9-fold lower affinity. ImImImPy-γ-PyPyPyPy-β-Dp binds the match site 5′-AGGGAA-3′ with an equilibrium association constant of $K_a = 3.7 \times 10^8$ M^{-1}. The end mismatch 5′-TGGGTC-3′ is bound with 26-fold lower affinity, and the two core-mismatches, 5′-AGGCAA-3′ and 5′-TGGGCT-3′, are bound with 130-fold and 220-fold lower affinity, respectively. Neither polyamide shows any appreciable binding to the core mismatch 5′-AGGAA(A)-3′ (data not shown).

Discussion

The two hairpin polyamides recognize the targeted 5′-AGGGA(A)-3′ sequence, as determined by MPE·Fe(II) footprinting and affinity cleaving, demonstrating specific recognition by polyamides of sequences containing *three contiguous* G·C base pairs, 5′-GGG-3′. Affinity cleavage data indicate that each polyamide is binding in the minor groove in a single orientation, consistent with the hairpin binding model (Figures 1 and 5). The relative intensities of the cleavage patterns are consistent with quantitative DNase I footprint titration results.

Quantitative DNase I footprint titrations reveal that ImImIm-γ-PyPyPy-β-Dp binds the designed match sites 5′-AGGGA-3′ and 5′-TGGGT-3′ with equilibrium association constants of $K_a = 5 \times 10^6$ and 8×10^6 M^{-1}, respectively. For comparison, the analogous six-ring hairpins containing only one and two contiguous imidazoles, ImPyPy-γ-PyPyPy-β-Dp and ImImPy-γ-PyPyPy-β-Dp, bind their respective match sites with affinities of $K_a \approx 10^8$ M^{-1}.[8,9] The addition of the third imidazole reduces binding affinity significantly, perhaps due to the inability of the polyamide to sit deeply in the sterically crowded minor groove, decreasing energetically favorable van der Waals contacts. Examination of the eight-ring hairpin, ImImImPy-γ-PyPyPyPy-β-Dp, reveals a dramatically increased affinity, the match site 5′-AGGGAA-3′ being bound with an equilibrium association constant of $K_a = 4 \times 10^8$ M^{-1}. The 80-fold increase in affinity in expanding from a three-ring to a four-ring hairpin polyamide mirrors the 66-fold enhancement of ImPyPyPy-Dp over ImPyPy-Dp that was observed in the 2:1 homodimeric polyamide:DNA motif.[13]

In addition to the observation that the eight-ring ImImImPy-γ-PyPyPyPy-β-Dp binds with higher affinity than the six ring ImImIm-γ-PyPyPy-β-Dp, the enhanced specificity is perhaps more significant. The six-ring hairpin polyamide binds 5′-TGGGC-3′, an end mismatch site, with 6-fold lower affinity compared to 5′-TGGGT-3′ (the highest affinity match), while the eight-ring hairpin polyamide binds its end mismatch site 5′-TGGGTC-3′ with 26-fold lower affinity compared to its match site 5′-AGGGAA-3′. Similarly, 5′-AGGCA-3′, a site containing a single base pair core mismatch, is bound by the six-ring system with 9-fold lower affinity, while the two core mismatch sites for the eight-ring system, 5′-AGGCAA-3′ and 5′-TGGGCT-3′, are recognized with 130-fold and 220-fold lower affinity, respectively. For both polyamides, binding of a site with a core single base pair mismatch results in a greater energetic penalty than binding of a site with single base pair mismatch in the end position. The increased specificity of ImImImPy-γ-PyPyPyPy-β-Dp over ImImIm-γ-PyPyPy-β-Dp may indicate that an imidazole placed immediately to the N-terminal side of the γ turn does not form as strong a hydrogen bond as in other positions.

Implications for the Design of Minor Groove Binding Molecules. Pyrrole–imidazole polyamides have been used to recognize a variety of target sequences containing A·T and G·C base pairs.[1,2,4,5,9] By recognizing sequences containing *three* contiguous G·C base pairs, 5′-(A,T)GGG(A,T)-3′ and 5′-(A,T)-GGG(A,T)$_2$-3′, this work expands the sequence-composition repertoire targetable by the hairpin polyamide motif. Both affinity and specificity for a G,C rich sequence are increased by the use of an eight-ring hairpin polyamide. This ability to enlarge the sequence repertoire, combined with rapid solid-phase synthesis, brings us one step closer to a universal approach for the recognition of any desired DNA sequence by strictly chemical methods.

Experimental Section

Dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBt), 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), and 0.2 mmol/g of Boc-β-alanine–(4-carboxamidomethyl) benzyl ester–copoly(styrene–divinylbenzene) resin (Boc-β-Pam-Resin) were purchased from Peptides International. *N,N*-Diiso-

(13) Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* 1996, *93*, 6981.

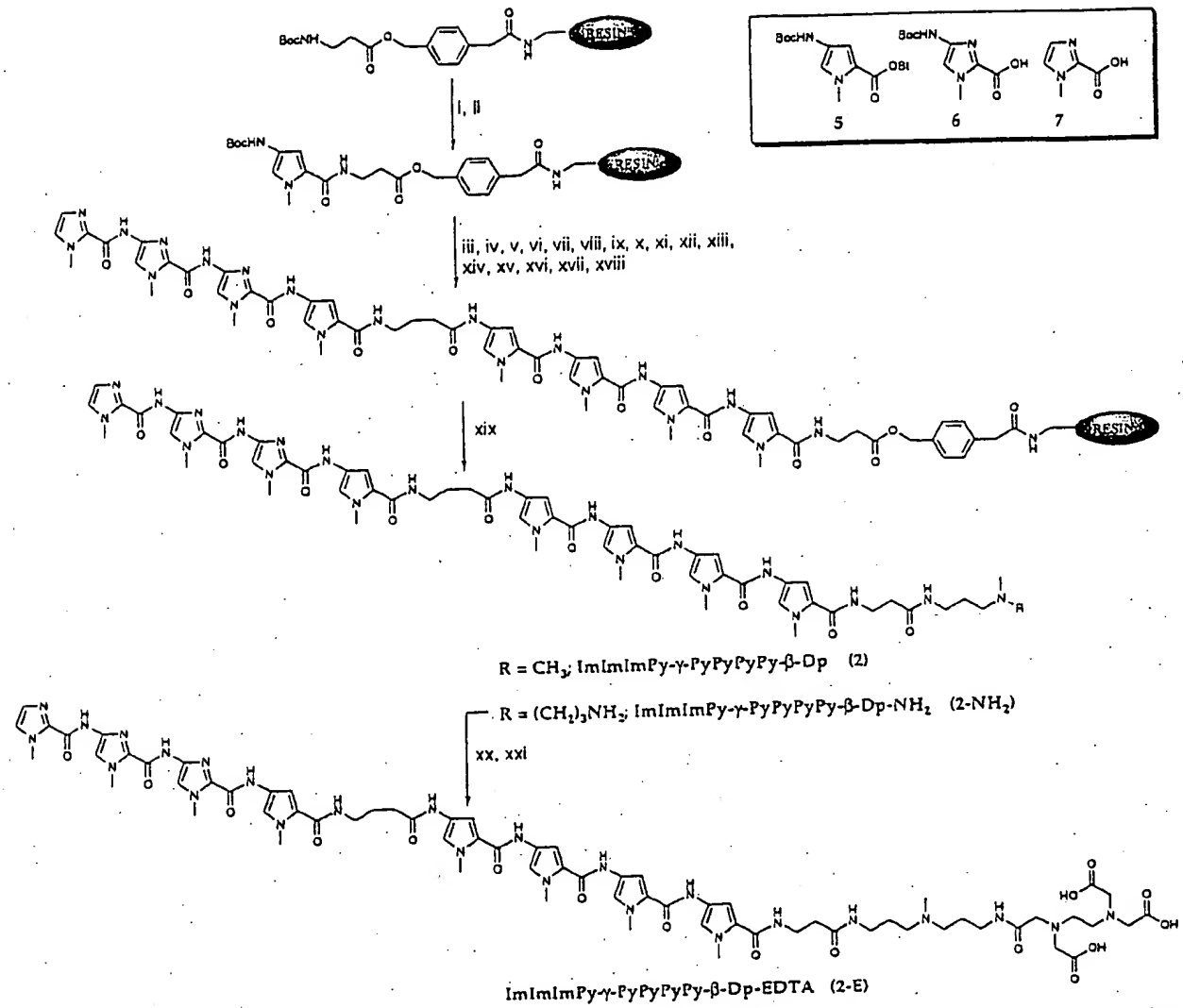


Figure 3. (Box) Pyrrole and imidazole monomers for synthesis of all compounds described here: Boc-pyrrole-OBt ester 5,[7] Boc-imidazole-acid 6,[7] and imidazole-2-carboxylic acid 7.[18] Solid-phase synthetic scheme for ImImImPy-γ-PyPyPyPy-β-Dp, ImImImPy-γ-PyPyPyPy-β-Dp-NH_2, and ImImImPy-γ-PyPyPyPy-β-Dp-EDTA prepared from commercially available Boc-β-alanine-Pam-resin (0.2 mmol/g): (i) 80% TFA/DCM, 0.4 M PhSH; (ii) BocPy-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4 M PhSH; (iv) BocPy-OBt, DIEA, DMF; (v) 80% TFA/DCM, 0.4 M PhSH; (vi) BocPy-OBt, DIEA, DMF; (vii) 80% TFA/DCM, 0.4 M PhSH; (viii) BocPy-OBt, DIEA, DMF; (ix) 80% TFA/DCM, 0.4 M PhSH; (x) Boc-γ-aminobutyric acid (HBTU, DIEA), DMF; (xi) 80% TFA/DCM, 0.4 M PhSH;3233 (xii) BocPy-OBt, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4 M PhSH; (xiv) BocIm-OBt (DCC/HOBt), DIEA, DMF; (xv) 80% TFA/DCM, 0.4 M PhSH; (xvi) BocIm-OBt (DCC/HOBt), DIEA, DMF; (xvii) 80% TFA/DCM, 0.4 M PhSH; (xviii) imidazole-2-carboxylic acid (HBTU/DIEA); (xix) ((*N*,*N*-dimethylamino)propyl)amine or 3,3'-diamino-*N*-methyldipropylamine, 55 °C; (xx) EDTA-dianhydride, DMSO/NMP, DIEA, 55 °C; (xxi) 0.1 M NaOH.

propylethylamine (DIEA), *N*,*N*-dimethylformamide (DMF), *N*-methylpyrrolidone (NMP), DMSO/NMP, acetic anhydride (Ac_2O), and 0.0002 M potassium cyanide/pyridine were purchased from Applied Biosystems. Boc-γ-aminobutyric acid was from NOVA Biochem, dichloromethane (DCM) and triethylamine (TEA) were reagent grade from EM, thiophenol (PhSH), ((dimethylamino)propyl)amine was from Aldrich, trifluoroacetic acid (TFA) was from Halocarbon, phenol was from Fisher, and ninhydrin was from Pierce. All reagents were used without further purification.

Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. and were used for filtration of DCU. Disposable polypropylene filters were also used for washing resin for ninhydrin and picric acid tests and for filtering predissolved amino acids into reaction vessels. A shaker for manual solid-phase synthesis was obtained from St. John Associates, Inc. Screw-cap glass peptide synthesis reaction vessels (5 and 20 mL) with a #2 sintered glass frit were made as described by Kent.[14] ^{1}H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz in DMSO-d_6, with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on either a HP 1090M analytical HPLC or a Beckman Gold system using a RAINEN C_{18}, Microsorb MV, 5 μm, 300 x 4.6 mm reversed phase column in 0.1% (wt/v) TFA with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% acetonitrile/min. Preparatory reverse-phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100 μm C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. The 18MΩ water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μm filtered.

ImImIm-γ-PyPyPy-β-Dp (1). The product was synthesized by manual solid-phase protocols[7] and recovered as a white powder (2.4 mg, 4% recovery). UV λ_{max} 312 (48 500); ^{1}H NMR (DMSO-d_6) δ

(14) Kent, S. B. H. *Annu. Rev. Biochem.* 1988, *57*, 957.

Figure 4. MPE·Fe(II) footprinting and affinity cleavage experiments on a 3′-^{32}P-labeled 274 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pSES1. The 5′-AGGGA(A)-3′, 5′-AGGCA(A)-3′, 5′-TGGGT-(C)-3′, 5′-TGGGC(T)-3′, and 5′-AGGAA(A)-3′ sites are shown on the right side of the autoradiogram. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lanes 4 and 11, MPE·Fe(II) standard; lane 5, 10 μM ImImImPy-γ-PyPyPyPy-β-Dp; lanes 6 and 7, 10 nM and 100 nM ImImImPy-γ-PyPyPyPy-β-Dp-EDTA·Fe(II); lane 8, 10 μM ImImIm-γ-PyPyPy-β-Dp; lanes 9 and 10, 100 nM and 1 μM ImImIm-γ-PyPyPy-β-Dp-EDTA·Fe(II). All lanes contain 15 kcpm 3′-radiolabeled DNA. The control and MPE·Fe(II) lanes (1, 4, 5, 8, and 11) contain 25 mM Tris−acetate buffer (pH 7.0), 10 mM NaCl, and 100 μM/base pair calf thymus DNA. The affinity cleavage lanes (6, 7, 9, and 10) contain 25 mM Tris−acetate buffer (pH 7.0), 200 mM NaCl, and 50 μg/mL glycogen.

10.09 (s, 1 H), 9.89 (s, 1 H), 9.88 (s, 1 H), 9.83 (s, 1 H), 9.57 (s, 1 H), 9.19 (br s, 1 H), 8.36 (t, 1 H, J = 5.6 Hz), 8.03 (m, 2 H), 7.64 (s, 1 H), 7.51 (s, 1 H), 7.45 (s, 1 H), 7.20 (d, 1 H, J = 1.0 Hz), 7.15 (d, 1 H, J = 2.0 Hz), 7.14 (s, 1 H), 7.08 (s, 1 H), 7.04 (s, 1 H), 6.87 (d, 2 H, J = 2.2 Hz), 4.01 (s, 3 H), 3.99 (s, 3 H), 3.95 (s, 3 H), 3.82 (s, 3 H), 3.82 (s, 3 H), 3.79 (s, 3 H), 3.37 (q, 2 H, J = 5.8 Hz), 3.26 (q, 2 H, J = 6.1 Hz), 3.10 (q, 2 H, J = 6.1 Hz), 2.99 (m, 2 H), 2.73 (d, 6 H, J = 4.8 Hz), 2.34 (t, 2 H, J = 7.2 Hz), 2.27 (t, 2 H, J = 7.3 Hz), 1.79 (m, 4 H); MALDI-TOF-MS, 980.1 (980.1 calcd for M + H).

ImImImPy-γ-PyPyPyPy-β-Dp (2). The product was synthesized by manual solid-phase protocols[7] and recovered as a white powder (7.6 mg, 11% recovery). UV λ_{max} 248 (42 000), 312 (48 500); ^{1}H NMR (DMSO-d_6) δ 10.32 (s, 1 H), 10.13 (s, 1 H), 9.93 (s, 1 H), 9.90 (s, 1 H), 9.89 (s, 1 H), 9.84 (s, 1 H), 9.59 (s, 1 H), 9.23 (br s, 1 H), 8.09 (t, 1 H, J = 5.3 Hz), 8.04 (m, 2 H), 7.65 (s, 1 H), 7.57 (s, 1 H), 7.46 (d, 1 H, J = 0.6 Hz), 7.22 (m, 3 H), 7.16 (s, 2 H), 7.09 (d, 1 H, J = 0.8 Hz), 7.06 (d, 2 H, J = 1.1 Hz), 7.00 (d, 1 H, J = 1.7), 6.88 (d, 1 H, J = 1.8), 6.87 (d, 1 H, J = 1.8 Hz), 4.02 (s, 3 H), 4.00 (s, 3 H), 3.99 (s, 3 H), 3.84 (s, 3 H), 3.83 (s, 3 H), 3.83 (s, 3 H), 3.80 (s, 3 H), 3.79 (s, 3 H), 3.37 (q, 2 H, J = 6.2 Hz), 3.21 (q, 2 H, J = 6.4 Hz), 3.10 (q, 2 H, J = 6.2 Hz), 3.00 (m, 2 H), 2.73 (d, 6 H, J = 4.9 Hz), 2.34 (t, 2 H, J = 7.2 Hz), 2.28 (t, 2 H, J = 7.0 Hz), 1.76 (m, 4 H); MALDI-TOF-MS, 1225.9 (1224.3 calcd for M + H).

ImImIm-γ-PyPyPy-β-Dp-NH$_2$ (1-NH$_2$). A sample of machine-synthesized resin (350 mg, 0.17 mmol/g[15]) was placed in a 20 mL glass scintillation vial and treated with 2 mL of 3,3′-diamino-*N*-methyldipropylamine at 55 °C for 18 h. The resin was removed by filtration through a disposable propylene filter, and the resulting solution was dissolved in water to a total volume of 8 mL and purified directly by preparatory reversed-phase HPLC to provide ImImIm-γ-PyPyPy-β-Dp-NH$_2$ (28 mg, 41% recovery) as a white powder. ^{1}H NMR (DMSO-d_6) δ 10.14 (s, 1 H), 9.89 (s, 1 H), 9.88 (s, 1 H), 9.83 (s, 1 H), 9.6 (br s, 1 H), 9.59 (s, 1 H), 8.36 (t, 1 H, J = 5.5 Hz), 8.09 (t, 1 H, J = 5.0 Hz), 8.03 (t, 1 H, J = 5.0 Hz), 7.9 (br s, 3 H), 7.63 (s, 1 H), 7.50 (s, 1 H), 7.44 (s, 1 H), 7.19 (d, 1 H, J = 1.2 Hz), 7.13 (m, 2 H), 7.08 (d, 1 H, J = 1.3 Hz), 7.02 (d, 1 H, J = 1.2 Hz), 6.85 (m, 2 H), 4.01 (s, 3 H), 3.99 (s, 3 H), 3.97 (m, 6 H), 3.80 (s, 3 H), 3.77 (s, 3 H), 3.34 (q, 2 H, J = 5.3 Hz), 3.23 (q, 2 H, J = 6.0 Hz), 3.05 (m, 6 H), 2.83 (q, 2 H, J = 5.0 Hz), 2.70 (d, 3 H, J = 4.0 Hz), 2.32 (t, 2 H, J = 6.9 Hz), 2.25 (t, 2 H, J = 6.9 Hz), 1.90 (m, 2 H), 1.77 (m, 4 H); MALDI-TOF-MS, 1022.8 (1023.1 calcd for M + H).

ImImImPy-γ-PyPyPyPy-β-Dp-NH$_2$ (2-NH$_2$). A sample of machine-synthesized resin (350 mg, 0.16 mmol/g[15]) was placed in a 20 mL glass scintillation vial and treated with 2 mL of 3,3′-diamino-*N*-methyldipropylamine at 55 °C for 18 h. The resin was removed by filtration through a disposable propylene filter, and the resulting solution was dissolved in water to a total volume of 8 mL and purified directly by preparatory reversed-phase HPLC to provide ImImImPy-γ-PyPyPyPy-β-Dp-NH$_2$ (31 mg, 40% recovery) as a white powder. ^{1}H NMR (DMSO-d_6) δ 10.37 (s, 1 H), 10.16 (s, 1 H), 9.95 (s, 1 H), 9.93 (s, 1 H), 9.91 (s, 1 H), 9.86 (s, 1 H), 9.49 (br s, 1 H), 9.47 (s, 1 H), 8.12 (m, 3 H), 8.0 (br s, 3 H), 7.65 (s, 1 H), 7.57 (s, 1 H), 7.46 (s, 1 H), 7.20 (m, 3 H), 7.16 (m, 2 H), 7.09 (d, 1 H, J = 1.5 Hz), 7.05 (m, 2 H), 7:00 (d, 1 H, J = 1.6 Hz), 6.88 (m, 2 H), 4.01 (s, 3 H), 3.99 (s, 3 H), 3.98 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.36 (q, 2 H, J = 5.3 Hz), 3.21−3.05 (m, 8 H), 2.85 (q, 2 H, J = 4.9 Hz), 2.71 (d, 3 H, J = 4.4 Hz), 2.34 (t, 2 H, J = 5.9 Hz), 2.26 (t, 2 H, J = 5.9 Hz), 1.85 (quintet, J = 5.7 Hz), 1.72 (m, 4 H); MALDI-TOF-MS, 1267.1 (1267.4 calcd for M + H).

ImImIm-γ-PyPyPy-β-Dp-EDTA (1-E). EDTA dianhydride (50 mg) was dissolved in 1 mL of DMSO/NMP solution and 1 mL of DIEA by heating at 55 °C for 5 min. The dianhydride solution was added to ImImIm-γ-PyPyPy-β-Dp-NH$_2$ (1-NH$_2$) (8.0 mg, 7 μmol) and dissolved in 750 μL of DMSO. The mixture was heated at 55 °C for 25 min, treated with 3 mL of 0.1 M NaOH, and heated at 55 °C for 10 min. TFA (0.1%) was added to adjust the total volume to 8 mL and the solution was purified directly by preparatory HPLC chromatography to provide 1-E as a white powder (3.3 mg, 30% recovery). ^{1}H NMR (DMSO-d_6) δ 10.14 (s, 1 H), 9.90 (s, 1 H), 9.89 (s, 1 H), 9.85 (s, 1 H), 9.58 (s, 1 H), 9.3 (br s, 1 H), 8.40 (m, 2 H), 8.02 (m, 2 H), 7.65 (s, 1 H), 7.51 (s, 1 H), 7.45 (s, 1 H), 7.20 (d, 1 H, J = 1.5 Hz), 7.15 (m, 2 H), 7.08 (d, 1 H, J = 1.1 Hz), 7.04 (d, 1 H, J = 1.5 Hz), 6.86 (m, 2 H), 4.00 (s, 3 H), 3.98 (s, 3 H), 3.94 (s, 3 H), 3.87 (m, 4 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.78 (s, 3 H), 3.72 (m, 4 H), 3.4−3.0 (m, 16 H), 2.71 (d, 3 H, J = 4.2 Hz), 2.33 (t, 2 H, J = 5.1 Hz), 2.25 (t, 2 H, J = 5.9 Hz), 1.75 (m, 6 H); MALDI-TOF-MS, 1298.4 (1298.3 calcd for M + H).

(15) Resin substitution has been corrected for the weight of the polyamide chain. The change in substitution during a specific coupling or for the entire synthesis can be calculated as L_{new}(mmol/g) = $L_{old}/(1 + L_{old}(W_{new} - W_{old}) \times 10^{-3})$, where L is the loading (mmol of amine per gram of resin), and W is the weight (g mol^{-1}) of the growing polyamide attached to the resin. See: Barlos, K.; Chatzi, O.; Gatos, D.; Stravropoulos, G. *Int. J. Peptide Protein Res.* 1991, *37*, 513.

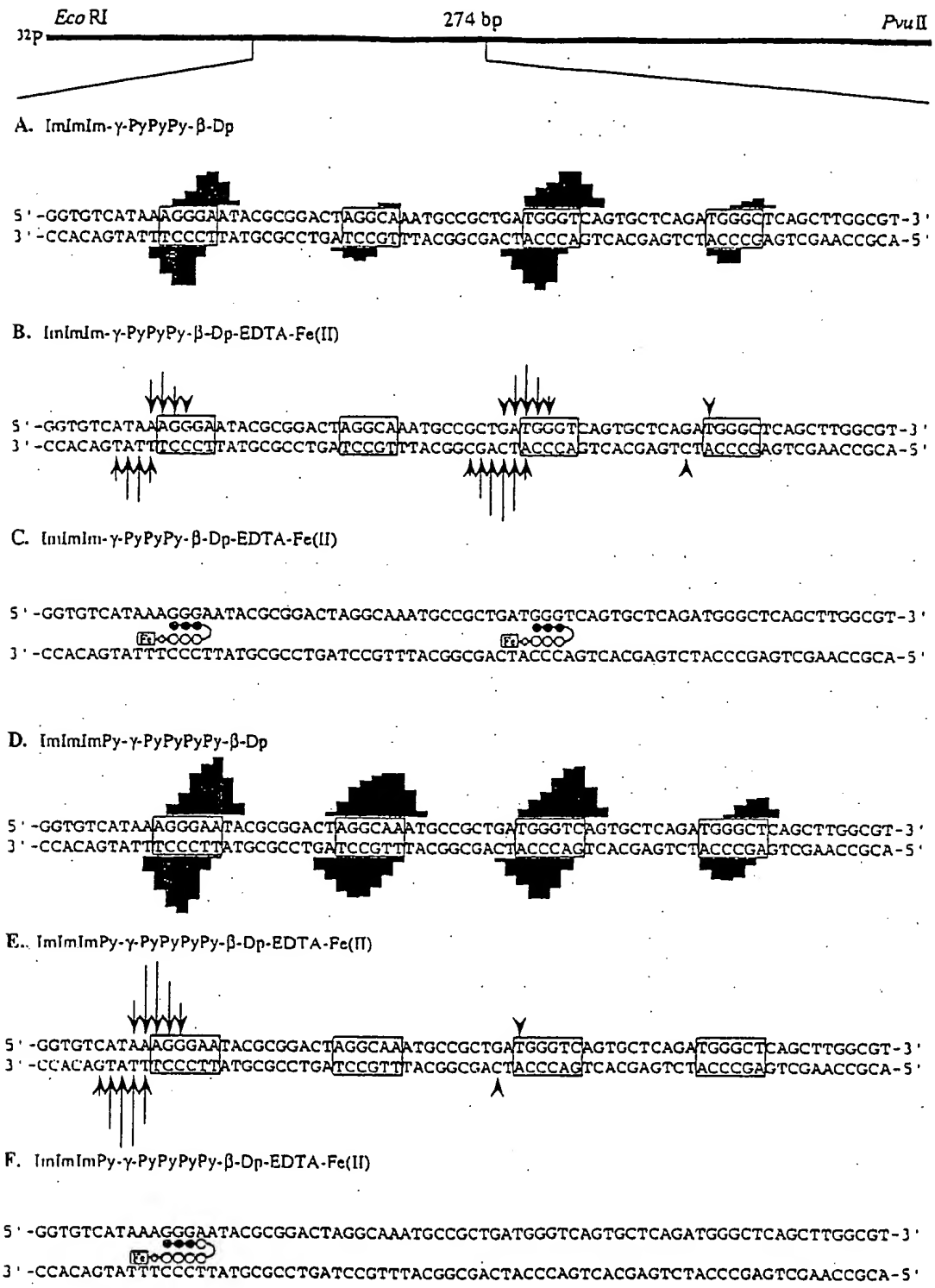


Figure 5. Results from MPE·Fe(II) footprinting of ImImIm-γ-PyPyPy-β-Dp and ImImImPy-γ-PyPyPyPy-β-Dp and affinity cleavage of ImImIm-γ-PyPyPy-β-Dp-EDTA·Fe(II) and ImImImPy-γ-PyPyPyPy-β-Dp-EDTA·Fe(II). (Top) Illustration of the 274 bp restriction fragment with the position of the sequence indicated. Boxes represent equilibrium binding sites determined by the published model. Only sites that were quantitated by DNase I footprint titrations are boxed. (A and D): MPE·Fe(II) protection patterns for polyamides at 10 μM concentration. Bar heights are proportional to the relative protection from cleavage at each band. (B and E): Affinity cleavage patterns of ImImIm-γ-PyPyPy-β-Dp-EDTA·Fe(II) at 1 μM and of ImImImPy-γ-PyPyPyPy-β-Dp-EDTA·Fe(II) at 100 nM, respectively. Arrow heights are proportional to the relative cleavage intensities at each base pair. (C and F): Ball and stick binding models for the single orientation binding to formal match sequences by the six-ring and eight-ring EDTA·Fe(II) analogs, respectively. Shaded and unshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent the β-alanine residue. The boxed Fe denotes the EDTA·Fe(II) cleavage moiety.

ImImImPy-γ-PyPyPyPy-β-Dp-EDTA (2-E). Compound 2-E was prepared as described for compound 1-E (yield 3.8 mg, 40%). ^{1}H NMR (DMSO-d_6) δ 10.34 (s, 1 H), 10.11 (s, 1 H), 9.92 (s, 1 H), 9.90 (s, 1 H), 9.89 (s, 1 H), 9.84 (s, 1 H), 9.57 (s, 1 H), 8.42 (m, 1 H), 8.03 (m, 3 H), 7.64 (s, 1 H), 7.56 (s, 1 H), 7.44 (s, 1 H), 7.20 (m, 3 H), 7.15 (m, 2 H), 7.07 (d, 1 H, J = 1.6 Hz), 7.05 (m, 2 H), 6.99 (d, 1 H, J = 1.6

Table 1. Equilibrium Association Constants (M^{-1})[a,b,c]

	match site		end mismatch	core mismatches
polyamide	5′-AGGGA-3′	5′-TGGGT-3′	5′-TGGGC-3′	5′-AGGCA-3′
ImImIm-γ-PyPyPy-β-Dp	4.6×10^6 (0.3)	7.6×10^6 (0.5)	1.3×10^6 (0.3)	8.6×10^5 (0.4)
	match site	end mismatch	core mismatches	
polyamide	5′-AGGGAA-3′	5′-TGGGTC-3′	5′-TGGGCT-3′	5′-AGGCAA-3′
ImImImPy-γ-PyPyPyPy-β-Dp	3.7×10^8 (0.8)	1.4×10^7 (0.5)	1.7×10^6 (0.3)	2.9×10^6 (0.3)

[a] Values reported are the mean values measured from a minimum of three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. [b] The assays were performed at 22 °C at pH 7.0 in the presence of 10 mM Tris−HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [c] Base pairs that are in bold represent formal mismatches.

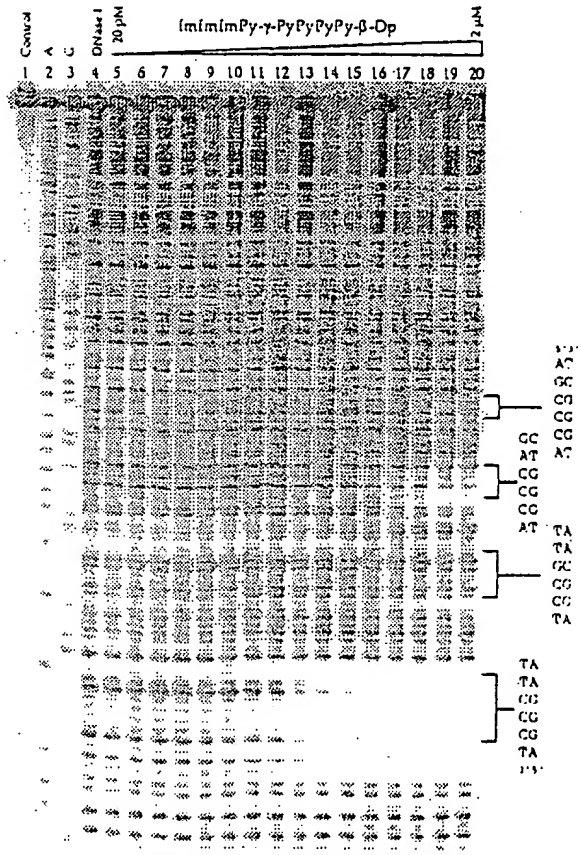


Figure 6. Quantitative DNase I footprint titration experiment with ImImImPy-γ-PyPyPyPy-β-Dp on the *Eco*RI/*Pvu*II restriction fragment from plasmid pSES1: lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4, DNase I standard; lanes 5−20, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM, 2 μM ImImImPy-γ-PyPyPyPy-β-Dp. The 5′-AGGGAA-3′, 5′-AGGCAA-3′, 5′-TGGGTC-3′, and 5′-TGGGCT-3′ sites that were analyzed are shown on the right side of the autoradiogram. All reactions contain 20 kcpm restriction fragment, 10 mM Tris·HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

Hz), 6.87 (m, 2 H), 4.00 (s, 3 H), 3.98 (s, 3 H), 3.97 (s, 3 H), 3.83 (m, 4 H), 3.82 (s, 6 H), 3.79 (s, 3 H), 3.78 (s, 6 H), 3.67 (m, 4 H), 3.4−3.0 (m, 16 H), 2.71 (d, 3 H, J = 4.2 Hz), 2.34 (t, 2 H, J = 5.4 Hz), 2.25 (t, 2 H, J = 5.9 Hz), 1.72 (m, 6 H); MALDI-TOF-MS, 1542.2 (1542.6 calcd for M + H).

DNA Reagents and Materials. Enzymes were purchased from Boehringer-Mannheim or New England Biolabs and were used with their supplied buffers. Deoxyadenosine and thymidine 5′-[α-^{32}P]-triphosphates and deoxyadenosine 5′-[γ-^{32}P]triphosphate were obtained from Amersham. Purified water was obtained by filtering doubly-distilled water through the MilliQ filtration system from Millipore. Sonicated, deproteinized calf thymus DNA was acquired from Pharmacia. All other reagents and materials were used as received. All DNA manipulations were performed according to standard protocols.[16]

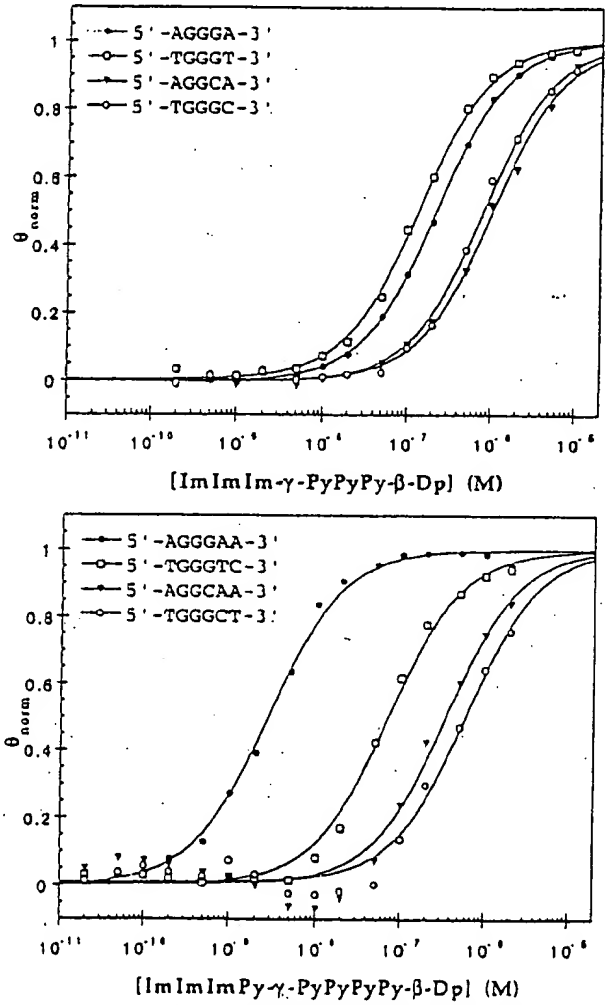


Figure 7. Data from the quantitative DNase I footprint titration experiments for the two polyamides, ImImIm-γ-PyPyPy-β-Dp (top) and ImImImPy-γ-PyPyPyPy-β-Dp (bottom), in complex with the designated sites. The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography and processed as described in the Experimental Section. The data points for 5′-AGGGA(A)-3′, 5′-TGGGT(C)-3′, 5′-AGGCA(A)-3′, and 5′-TGGGC(T)-3′ sites are indicated by filled circles (●), open squares (□), filled inverted triangles (▼), and open circles (○), respectively. The solid curves are the best-fit Langmuir binding titration isotherms obtained from the nonlinear least-squares algorithm using eq 2.

Construction of Plasmid DNA. The plasmid pSES1 was constructed by hybridization of the inserts, 5′-GATCCGGTGTCATAAAGGGAATACGCGGACTAGGCAAATGCCGC-TGATGGGTCAGTGCTCAGATGGGCTC-3′ and 5′-AGCTGAGC-

(16) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.

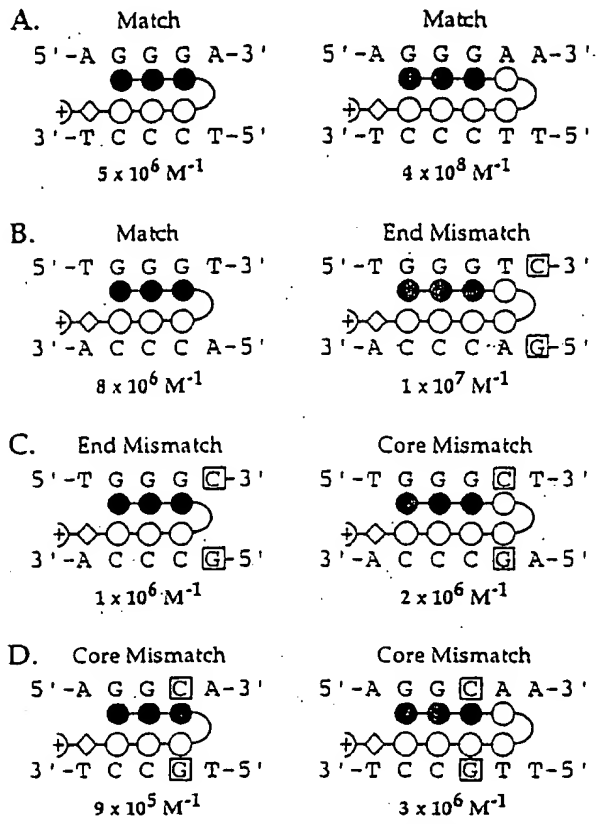


Figure 8. Ball and stick models of ImImIm-γ-PyPyPy-β-Dp (left) and ImImImPy-γ-PyPyPyPy-β-Dp (right) for each binding site, with the corresponding equilibrium association constants shown below each individual model. The binding sites shown are (a) 5'-AGGGA(A)-3', (b) 5'-TGGGT(C)-3', (c) 5'-TGGGC(T)-3', and (d) 5'-AGGCA(A)-3'. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent the β-alanine residue. Formally mismatched base pairs are boxed.

CCATCTGAGCACTGACCCATCAGCGGCATTTGCCTAGT-CCGCGTATTCCCTTTATGACACCG-3'. The hybridized insert was ligated into linearized pUC19 *Bam*HI/*Hin*dIII plasmid using T4 DNA ligase. The resultant constructs were used to transform Epicurian Coli XL-2 Blue competent cells from Stratagene. Ampicillin-resistant white colonies were selected from 25 mL of Luria-Bertani medium agar plates containing 50 μg/mL ampicillin and treated with XGAL and IPTG solutions. Large-scale plasmid purification was performed with Qiagen Maxi purification kits. Dideoxy sequencing was used to verify the presence of the desired insert. Concentration of the prepared plasmid was determined at 260 nm using the relationship of 1 OD unit = 50 μg/mL of duplex DNA.

Preparation of 3'- and 5'-End-Labeled Restriction Fragments. The plasmid pSES1 was linearized with *Eco*RI and then treated with either Klenow fragment, deoxyadenosine 5'-[α-^{32}P]triphosphate and thymidine 5'-[α-^{32}P]triphosphate for 3' labeling, or calf alkaline phosphatase and then 5' labeled with T4 polynucleotide kinase and deoxyadenosine 5'-[γ-^{32}P]triphosphate. The labeled fragment (3' or 5') was then digested with *Pvu*II and loaded onto a 5% non-denaturing polyacrylamide gel. The desired 274 base pair band was visualized by autoradiography and isolated. Chemical sequencing reactions were performed according to published methods.[17]

MPE·Fe(II) Footprinting. All reactions were carried out in a volume of 40 μL. A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were 25 mM Tris–acetate buffer (pH 7.0), 10 mM NaCl, 100 μM/base pair calf thymus DNA, and 15 kcpm 3'- or 5'-radiolabeled DNA. The solutions were allowed to equilibrate for 4 h. A fresh 50 μM MPE·Fe(II) solution was made from 100 μL of a 100 μM MPE solution and 100 μL of a 100 μM ferrous ammonium sulfate ($Fe(NH_4)_2(SO_4)_2\cdot 6H_2O$) solution. After the 4 h equilibration, MPE·Fe(II) solution (5 μM) was added, and the reactions were equilibrated for 5 min. Cleavage was initiated by the addition of dithiothreitol (5 mM) and allowed to proceed for 14 min. Reactions were stopped by ethanol precipitation, resuspended in 100 mM tris-borate–EDTA/80% formamide loading buffer, denatured at 85 °C for 5 min, placed on ice, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V.

Affinity Cleaving. All reactions were carried out in a volume of 400 μL. A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were 25 mM Tris–acetate buffer (pH 7.0), 200 mM NaCl, 50 μg/mL of glycogen, and 15 kcpm 3'- or 5'-radiolabeled DNA. After the reactions were allowed to equilibrate for 4 h, ferrous ammonium sulfate ($Fe(NH_4)_2(SO_4)_2\cdot 6H_2O$), 10 μM final concentration, was added. After another 15 min, cleavage was initiated by the addition of dithiothreitol (5 mM) and allowed to proceed for 12 min. Reactions were stopped by ethanol precipitation, resuspended in 100 mM tris-borate–EDTA/80% formamide loading buffer, denatured at 85 °C for 5 min, placed on ice, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% cross-link, 7 M urea) at 2000 V.

DNase I Footprinting. All reactions were carried out in a volume of 40 μL. We note explicitly that no carrier DNA was used in these reactions. A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were 10 mM Tris·HCl buffer (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, and 20 kcpm 3'-radiolabeled DNA. The solutions were allowed to equilibrate for a minimum of 4 h at 22 °C (the four-ring hairpin was allowed to equilibrate for up to 12 h with no noticeable affect on the data set). Cleavage was initiated by the addition of 4 μL of a DNase I stock solution (diluted with 1 mM DTT to give a stock concentration of 0.225 u/mL) and was allowed to proceed for 5 min at 22 °C. The reactions were stopped by the addition of 3 M sodium acetate solution containing 50 mM EDTA and then ethanol precipitated. The cleavage products were resuspended in 100 mM tris-borate–EDTA/80% formamide loading buffer, denatured at 85 °C for 5 min, placed on ice, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% cross-link, 7 M urea) at 2000 V for 1 h. The gels were dried under vacuum at 80 °C, then quantitated using storage phosphor technology.

Equilibrium association constants were determined as previously described.[6,11] The data were analyzed by performing volume integrations of the 5'-AGGGA(A)-3', 5'-TGGGT(C)-3', 5'-TGGGC(T)-3', and 5'-AGGGA(A)-3' sites and a reference site. The apparent DNA target site saturation, θ_{app}, was calculated for each concentration of polyamide using the following equation:

$$\theta_{app} = 1 - \frac{I_{tot}/I_{ref}}{I_{tot}^{\circ}/I_{ref}^{\circ}} \qquad (1)$$

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{ref}° correspond to those values for a DNase I control lane to which no polyamide has been added. The ($[L]_{tot}$, θ_{app}) data points were fit to a Langmuir binding isotherm (eq 2, $n = 1$) by minimizing the difference between θ_{app} and θ_{fit}, using the modified Hill equation:

$$\theta_{fit} = \theta_{min} + (\theta_{max} - \theta_{min})\frac{K_a^n[L]_{tot}^n}{1 + K_a^n[L]_{tot}^n} \qquad (2)$$

where $[L]_{tot}$ corresponds to the total polyamide concentration, K_a corresponds to the apparent monomeric association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software (version 2.1, Abelbeck software) with K_a, θ_{max}, and θ_{min} as the adjustable parameters. All acceptable fits had a correlation coefficient of $R > 0.97$. At least three sets of acceptable data were used in determining each association constant. All lanes from each gel were used unless visual inspection revealed a data point to be

(17) (a) Iverson, B. L.; Dervan, P. B. *Nucleic Acids Res.* **1987**, *15*, 7823. (b) Maxam, A. M.; Gilbert, W. S. *Methods Enzymol.* **1980**, *65*, 499.

obviously flawed relative to neighboring points. The data were normalized using the following equation:

$$\theta_{norm} = \frac{\theta_{app} - \theta_{min}}{\theta_{max} - \theta_{min}} \quad (3)$$

Quantitation by Storage Phosphor Technology Autoradiography. Photostimulable storage phosphorimaging plates (Kodak Storage Phosphor Screen S0230 obtained from Molecular Dynamics) were pressed flat against gel samples and exposed in the dark at 22 °C for 12−16 h. A Molecular Dynamics 400S PhosphorImager was used to obtain all data from the storage screens. The data were analyzed by performing volume integrations of all bands using the ImageQuant v. 3.2.

Acknowledgment. This paper is dedicated to Nelson Leonard on the occasion of his 80th birthday. We are grateful to the National Institutes of Health (GM-2768) for research support, the National Institutes of Health for a research service award to S.E.S., and the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B. We would like to thank G. M. Hathaway for MALDI-TOF mass spectrometry.

JA9611598

Recognition of 5′-(A,T)GG(A,T)$_2$-3′ Sequences in the Minor Groove of DNA by Hairpin Polyamides

Michelle E. Parks, Eldon E. Baird, and Peter B. Dervan*

Contribution from the Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125

Received March 6, 1996[⊗]

Abstract: A series of four hairpin pyrrole–imidazole polyamides, ImImPy-γ-PyPyPy-β-Dp, PyPyPy-γ-ImImPy-β-Dp, AcImImPy-γ-PyPyPy-β-Dp, and AcPyPyPy-γ-ImImPy-β-Dp (Im = *N*-methylimidazole-2-carboxamide, Py = *N*-methylpyrrole-2-carboxamide, Dp = *N*,*N*-dimethylaminopropylamide, γ = γ-aminobutyric acid, β = β-alanine, and Ac = acetyl), designed for recognition of 5′-(A,T)GG(A,T)$_2$-3′ sequences in the minor groove of DNA were synthesized using solid phase methodology and analyzed with respect to DNA binding affinity and sequence specificity. Quantitative DNase I footprint titration experiments reveal that the optimal polyamide ImImPy-γ-PyPyPy-β-Dp binds a designated 5′-TGGTT-3′ match site with an equilibrium association constant of $K_a = 1.0 \times 10^8$ M^{-1} and the single base pair mismatch sites, 5′-TGTTA-3′ and 5′-GGGTA-3′, with 50-fold and 100-fold-lower affinity, respectively (10 mM Tris·HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0 and 22 °C). Polyamides of sequence composition AcImImPy-γ-PyPyPy-β-Dp and AcPyPyPy-γ-ImImPy-β-Dp, which differ only by the position of the γ-linker, bind with similar affinities and specificities. Recognition of sequences containing contiguous G·C base pairs expands the sequence repertoire available for targeting DNA with pyrrole–imidazole polyamides.

Introduction

Pyrrole–imidazole polyamides offer a general method for the design of non-natural molecules for sequence-specific recognition in the minor groove of DNA.[1-3] Within the 2:1 polyamide–DNA model, an imidazole (Im) on one ligand opposite a pyrrolecarboxamide (Py) on the second ligand recognizes a G·C base pair, while a pyrrolecarboxamide/imidazole combination targets a C·G base pair.[1,3] A pyrrolecarboxamide/pyrrolecarboxamide pair is partially degenerate for A·T or T·A base pairs.[1-3] On the basis of this model, the recognition of the sequences 5′-(A,T)G(A,T)C(A,T)-3′,[1] 5′-(A,T)G(A,T)$_3$-3′,[3] (A,T)$_2$G(A,T)$_2$-3′,[4] and 5′-(A,T)GCGC(A,T)-3′[5] has been achieved. However, sequences containing *contiguous* G·C base pairs are notably absent from this list.

Formation of a hairpin polyamide by covalently linking a polyamide heterodimer with a γ-aminobutyric acid (γ) residue provides an approximate 300-fold enhancement in affinity over the unlinked polyamides, ImPyPy- and PyPyPy-.[6] Moreover, the specificity of the hairpin is greatly improved. The initial placement of the γ-amino acid turn was chosen for synthetic ease and was not varied. With the development of solid phase methodology for polyamide synthesis, we now assess the effect of varying the position of the γ-turn monomer.[7]

In order to explore the recognition of 5′-(A,T)GG(A,T)$_2$-3′ sequences, a series of four head-to-tail linked hairpin polyamides containing *neighboring* imidazole rings, ImImPy-γ-PyPyPy-β-Dp (1), PyPyPy-γ-ImImPy-β-Dp (2), AcImImPy-γ-PyPyPy-β-Dp (3), and AcPyPyPy-γ-ImImPy-β-Dp (4), were prepared using solid phase methods (Figures 1 and 2).[7] The polyamides are all synthesized with Boc-β-alanine-Pam-resin, previously shown as optimal for polyamides.[8] Each imidazole is expected to form a specific hydrogen bond with a guanine amino group allowing the recognition of contiguous G·C base pairs (Figure 1). In addition, the linker turn position is varied within the nonacetylated and acetylated pairs of polyamides to determine the effect on the sequence specificity and binding affinity. We report here the binding specificity and affinity of the polyamides as determined by the complementary techniques, MPE·Fe^{II} footprinting[9] and quantitative DNase I footprinting.[10] MPE·Fe^{II} footprinting verifies that sequence-specific recognition of the expected 5′-TGGTT-3′ target site has been achieved. In addition, DNase I quantitative footprint titration experiments reveal that the position of the γ-linker does not dramatically affect either affinity or specificity of polyamides, especially the pair containing acetylated N-termini.

Results

Synthesis of Polyamides. The polyamides ImImPy-γ-PyPyPy-β-Dp (1), PyPyPy-γ-ImImPy-β-Dp (2), AcImImPy-γ-PyPyPy-β-Dp (3), and AcPyPyPy-γ-ImImPy-β-Dp (4) were

[⊗] Abstract published in *Advance ACS Abstracts,* June 15, 1996.

(1) (a) Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574–1575. (b) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (c) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci., U.S.A.* **1992**, *89*, 7586. (d) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385.

(2) (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingham, M. *Nature Struct. Biol.* **1994**, *1*, 169.

(3) (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (b) Geierstanger, B. H.; Jacobsen, J.-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994**, *33*, 3055.

(4) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.

(5) (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646–650. (b) Mrksich, M.; Dervan, P. B.; *J. Am. Chem. Soc.* **1995**, *117*, 3325.

(6) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983.

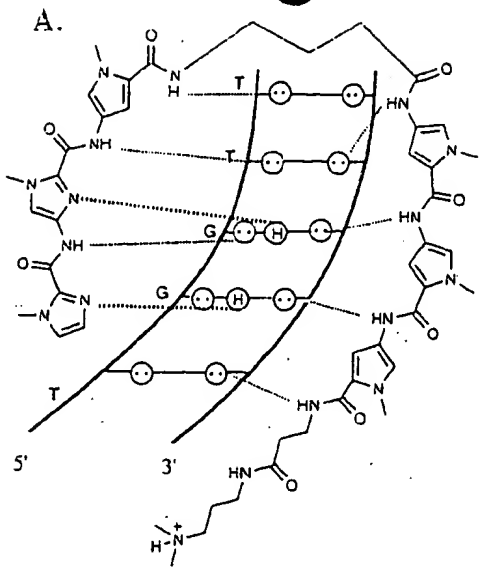
(7) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141–6146.

(8) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147–6152.

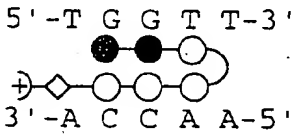
(9) (a) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373. (b) Van Dyke, M. W.; Dervan, P. B. *Nucleic Acids Res.* **1983**, *11*, 5555.

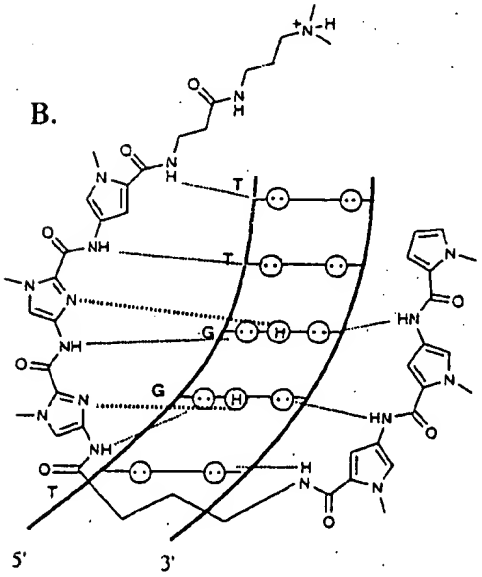
(10) (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8462. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.

S0002-7863(96)00728-7 CCC: $12.00



ImImPy-γ-PyPyPy-β-Dp • TGGTT





PyPyPy-γ-ImImPy-β-Dp • TGGTT

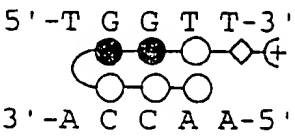
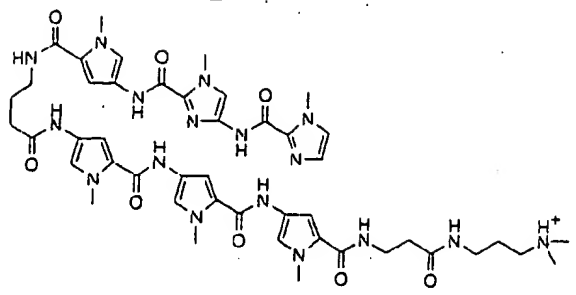
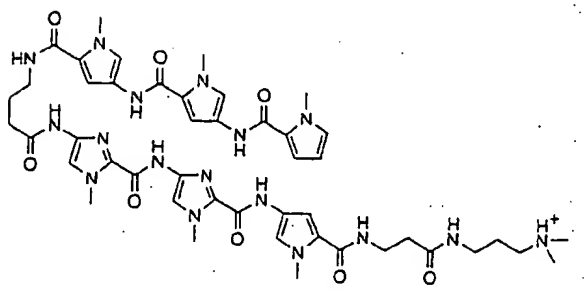


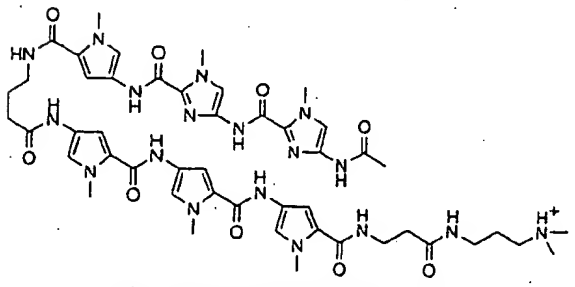
Figure 1. Binding model for the complexes formed between polyamides ImImPy-γ-PyPyPy-β-Dp (1) (A) and PyPyPy-γ-ImImPy-β-Dp (2) (B) and a 5′-TGGTT-3′ sequence. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. Ball and stick models are also shown. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent a β-alanine residue.



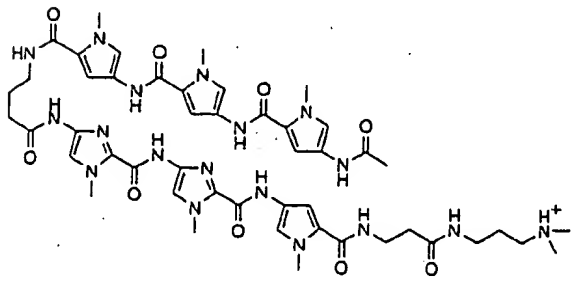
ImImPy-γ-PyPyPy-β-Dp (1)



PyPyPy-γ-ImImPy-β-Dp (2)



AcImImPy-γ-PyPyPy-β-Dp (3)



AcPyPyPy-γ-ImImPy-β-Dp (4)

Figure 2. Series of polyamides synthesized using solid phase methodology.[7]

prepared by solid phase methodology (Figure 2). Four unique pyrrole and imidazole building blocks were combined in a stepwise manner on a solid support using Boc-chemistry protocols (Figure 3). For example, polyamide 2, PyPyPy-γ-ImImPy-β-Dp, was prepared in 14 steps on the resin, and then cleaved with a single-step aminolysis reaction (Figure 3). All polyamides were found to be soluble up to at least 1 mM concentration in aqueous solution.

Footprinting. MPE·Fe^{II} footprinting on a 3′- or 5′-^{32}P-end-labeled 266 base pair *Eco*RI/*Pvu*II restriction fragment from plasmid pMEPGG (25 mM Tris-acetate, 100 μM bp calf thymus DNA, 10 mM NaCl) reveals that the synthetic polyamides 1−4, at 10 μM concentration, bind the designated target site 5′-

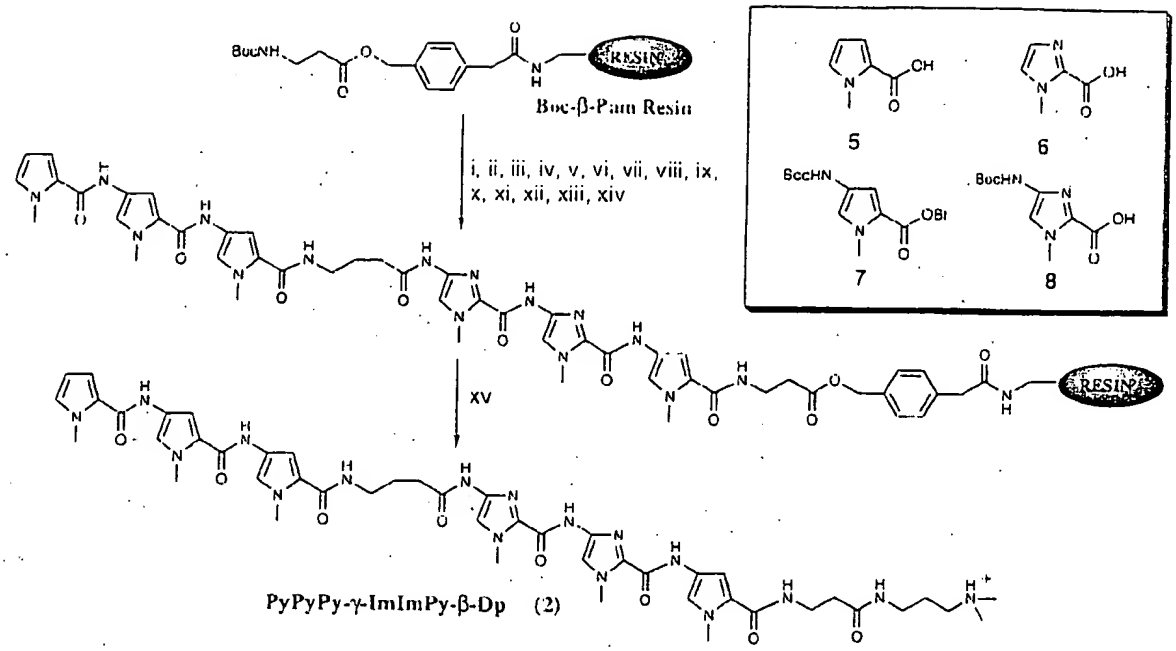


Figure 3. Solid phase synthetic scheme for PyPyPy-γ-ImImPy-β-Dp starting from commercially available Boc-β-Pam-resin: (i) 80% TFA/DCM, 0.4 M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4 M PhSH; (iv) Boc-Im-OBt (DCC/HOBt), DIEA, DMF; (v) 80% TFA/DCM, 0.4 M PhSH; (vi) Boc-Im-OBt (DCC/HOBt), DIEA, DMF; (vii) 80% TFA/DCM, 0.4 M PhSH; (viii) Boc-γ-aminobutyric acid (HBTU, DIEA); (ix) 80% TFA/DCM, 0.4 M PhSH; (x) Boc-Py-OBt, DIEA, DMF; (xi) 80% TFA/DCM, 0.4 M PhSH; (xii) Boc-Py-OBt, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4 M PhSH; (xiv): pyrrole-2-carboxylic acid (HBTU/DIEA); (xv) (*N,N*-dimethylamino)propylamine, 55 °C. (Inset) Pyrrole and imidazole monomers for synthesis of all compounds described here: pyrrole-2-carboxylic acid **5**, imidazole-2-carboxylic acid **6**,[1b] Boc−pyrrole−OBt ester **7**,[7] and Boc−imidazole acid **8**.[7]

TGGTT-3′ (Figures 4 and 5). In addition, several single base pair mismatch sites are bound with lower affinity. Quantitative DNase I footprint titration experiments (10 mM Tris·HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0 and 22 °C) were performed to determine the equilibrium association constants of the four polyamides **1**−**4** for a designated match site, 5′-TGGTT-3′, as well as for two single base pair mismatch sites, 5′-TGTTA-3′ and 5′-GGGTA-3′ (Table 1).[10] The polyamide ImImPy-γ-PyPyPy-β-Dp binds the target site 5′-TGGTT-3′ with the highest affinity (association constant K_a = 1.0 × 10^8 M^{-1}) (Figures 6 and 7). The remaining polyamides have lower but approximately equal association constants of K_a = ~2 × 10^7 M^{-1} for the target site. The nonacetylated polyamides in this series are >50-fold specific for the 5′-TGGTT-3′ match site over either of the single base pair mismatch sites analyzed. The acetylated pair of polyamides exhibit lower sequence specificity for the analyzed sites.

Discussion

Each polyamide within this series specifically binds the five base pair designated target sequence 5′-TGGTT-3′, as shown by MPE·Fe^{II} footprinting experiments, providing the first example of contiguous G·C recognition in the polyamide−DNA motif. Interestingly, the polyamides prefer different mismatch sequences, indicating that the position of the turn alters sequence selectivity, although only for the mismatches.

Quantitative DNase I footprint titration experiments reveal that ImImPy-γ-PyPyPy-β-Dp (**1**) is optimal within this series of four polyamides. This hairpin binds a 5′-TGGTT-3′ match site with an equilibrium association constant of K_a = 1 × 10^8 M^{-1}, while the corresponding hairpin PyPyPy-γ-ImImPy-β-Dp (**2**), which differs only in the position of the γ turn, shows lower affinity (K_a = ~2 × 10^7 M^{-1}) for the 5′-TGGTT-3′ site. Both unacetylated polyamides demonstrate good specificity (>50-fold) for the target match site over the single base pair mismatch sites. The acetylated polyamides are similar in affinity to PyPyPy-γ-ImImPy-β-Dp (**2**), but exhibit lower specificity. AcImImPy-γ-PyPyPy-β-Dp (**3**) and AcPyPyPy-γ-ImImPy-β-Dp (**4**) are virtually indistinguishable from each other on the basis of affinity and specificity for the analyzed target sequences, indicating little preference for turn position.

This series of contiguous imidazole-containing polyamides is remarkably similar in affinity and specificity to the single imidazole-containing hairpin polyamide, ImPyPy-γ-PyPyPy-β-Dp, indicating little or no energetic penalty in this system for *adjacent* imidazoles.[8] Importantly, the position of the hairpin turn does not significantly affect the recognition of the target 5′-TGGTT-3′ match site, although single base pair mismatch relative affinities are altered.

Implications for the Design of Minor Groove Binding Molecules. The 2:1 motif has been used to specifically target several sequences: 5′-TGTCA-3′,[1] 5′-TGTTA-3′,[3] 5′-AAGTT-3′,[4] and 5′-TGCGCA-3′.[5] The results reported herein add sequences containing two contiguous G·C base pairs to the list, expanding the sequence repertoire for DNA recognition by polyamides. Furthermore, turn position showed minimal effects on the specificity and affinity of the polyamides, indicating a new degree of flexibility within the 2:1 motif. The expansion of the polyamide sequence repertoire through contiguous G·C recognition coupled with solid phase synthetic advances allowing the rapid assembly and characterization of polyamides brings the goal of sequence-specific recognition of any DNA sequence by designed molecules closer to fruition.

Experimental Section

Materials. Boc-glycine-(4-carbonylaminomethyl)-benzyl-ester-co-poly(styrene−divinylbenzene) resin (Boc-G-Pam-resin) (0.2 mmol/g) 0.2 mmol/g Boc-β-alanine-(4-carbonylaminomethyl)-benzyl-ester-co-

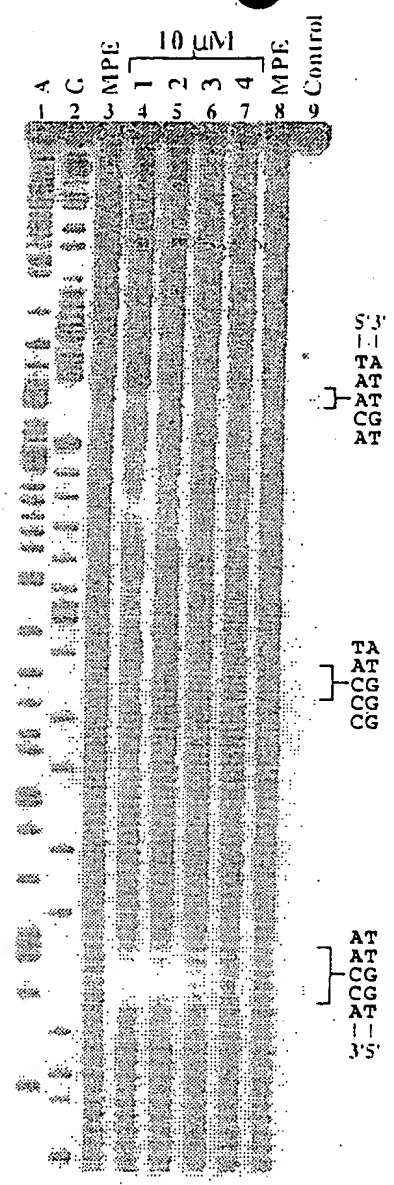


Figure 4. MPE·Fe^{II} footprinting experiment on a 3′-^{32}P-labeled 266 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pMEPGG. The 5′-TGGTT-3′, 5′-GGGTA-3′, and 5′-TGTTA-3′ sites are shown on the right side of the autoradiogram. All reactions contain 10 kcpm restriction fragment, 25 mM Tris-acetate, 10 mM NaCl, 100 μM calf thymus DNA (bp), and 5 mM DTT. Lane 1: A reaction. Lane 2: G reaction. Lanes 3 and 8: MPE·Fe^{II} standard. Lane 4: 10 μM ImImPy-γ-PyPyPy-β-Dp (1). Lane 5, 10 μM PyPyPy-γ-ImImPy-β-Dp (2). Lane 6: 10 μM AcImImPy-γ-PyPyPy-β-Dp (3). Lane 7: 10 μM AcPyPyPy-γ-ImImPy-β-Dp (4). Lane 9; intact DNA.

poly(styrene−divinylbenzene) resin (Boc-β-Pam-resin), dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBt), 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), Boc-glycine, and Boc-β-alanine were purchased from Peptides International. *N,N*-Diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF), *N*-methylpyrrolidone (NMP), DMSO/NMP, and acetic anhydride (Ac_2O) were purchased from Applied Biosystems. Boc-γ-aminobutyric acid was from NOVA Biochem, dichloromethane (DCM) and triethylamine (TEA) were reagent grade from EM, thiophenol (PhSH) and (dimethylamino)propylamine were from Aldrich, trifluoroacetic acid (TFA) was from Halocarbon. All reagents were used without further purification.

^{1}H NMR spectra were recorded in DMSO-d_6 on a GE 300 instrument operating at 300 MHz. Chemical shifts are reported in parts per million relative to the solvent residual signal. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed either on a HP 1090M analytical HPLC or on a Beckman Gold system using a Rainin C_{18}, Microsorb MV, 5 μm, 300 × 4.6 mm reversed-phase column in 0.1% (w/v) TFA with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% acetonitrile/min. Preparatory HPLC was carried out on a Beckman HPLC using a Waters DeltaPak 25 × 100 mm, 100 μm C_{18} column equipped with a guard, 0.1% (w/v) TFA, 0.25% acetonitrile/min. 18MΩ water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μm filtered. Reagent-grade chemicals were used unless otherwise stated.

Activation of Boc-γ-aminobutyric Acid, Imidazole-2-carboxylic acid, and Pyrrole-2-carboxylic acid. The appropriate amino acid or acid (2 mmol) was dissolved in 2 mL of DMF. HBTU (720 mg, 1.9 mmol) was added followed by DIEA (1 mL) and the solution lightly shaken for at least 5 min.

Activation of Boc−Imidazole Acid. Boc−imidazole acid (257 mg, 1 mmol) and HOBt (135 mg, 1 mmol) were dissolved in 2 mL of DMF, DCC (202 mg, 1 mmol) was then added, and the solution was allowed to stand for at least 5 min.

Typical Manual Synthesis Protocol. PyPyPy-γ-ImImPy-β-Dp. Boc-β-Pam-resin (1.25 g, 0.25 mmol of amine) was shaken in DMF for 30 min and drained. The N-Boc group was removed by washing with DCM for 2 × 30 s, followed by a 1 min shake in 80% TFA/DCM/0.5 M PhSH, draining the reaction vessel, a brief 80% TFA/DCM/0.5 M PhSH wash, and 20 min shaking in 80% TFA/DCM/0.5 M PhSH solution. The resin was washed for 1 min with DCM and 30 s with DMF. A resin sample (8−10 mg) was taken for analysis. The resin was drained completely, Boc-pyrrole-OBt monomer (357 mg, 1 mmol) dissolved in 2 mL of DMF was added, followed by DIEA (1 mL), and the resin was shaken vigorously to make a slurry. The coupling was allowed to proceed for 45 min. A resin sample (8−10 mg) was taken after 40 min to check reaction progress. The reaction vessel was washed with DMF for 30 s and dichloromethane for 1 min to complete a single reaction cycle. Six additional cycles were performed, adding successively Boc-Im-OH (DCC/HOBt), Boc-Im-OH (DCC/HOBt), Boc-γ-aminobutyric acid (HBTU/DIEA), Boc-Py-OBt, Boc-Py-OBt, and pyrrole-2-carboxylic acid (HBTU/DIEA). The resin was washed with DMF, DCM, MeOH, and ethyl ether and then dried *in vacuo*. PyPyPy-γ-ImImPy-β-Pam-resin (180 mg, 29 μmol)[12] was weighed into a glass scintillation vial, 1.5 mL of (*N,N*-dimethylamino)-propylamine added, and the mixture heated at 55 °C for 18 h. The resin was removed by filtration through a disposable polypropylene filter and washed with 5 mL of water, the amine solution and the water washes were combined, the solution was loaded on a C_{18} preparatory HPLC column. The polyamide was then eluted in 100 min as a well-defined peak with a gradient of 0.25% acetonitrile/min. The polyamide was collected in four separate 8 mL fractions, and the purity of the individual fractions was verified by HPLC and ^{1}H NMR, to provide purified PyPyPy-γ-ImImPy-β-Dp (2) (11.2 mg, 39% recovery): UV λ_{max} 246 (31 100), 312 (51 200); ^{1}H NMR (DMSO-d_6) δ 10.30 (s, 1 H), 10.26 (s, 1 H), 9.88 (s, 1 H), 9.80 (s, 1 H), 9.30 (s, 1 H), 9.2 (br s, 1 H), 8.01 (m, 3 H), 7.82 (br s 1 H), 7.54 (s, 1 H), 7.52 (s, 1 H), 7.20 (d, 1 H, J = 1.3 Hz), 7.18 (d, 1 H, J = 1.2 Hz), 7.15 (d, 1 H, J = 1.3 Hz), 7.01 (d, 1 H, J = 1.4 Hz), 6.96 (d, 1 H, J = 1.4 Hz), 6.92 (d, 1 H, J = 1.8 Hz), 6.89 (m, 2 H), 6.03 (t, 1 H, J = 2.4 Hz), 3.97 (s, 3 H), 3.96 (s, 3 H), 3.85 (s, 3 H), 3.82 (s, 3 H), 3.78 (m, 6 H), 3.37 (m, 2 H), 3.20 (q, 2 H, J = 5.7 Hz), 3.08 (q, 2 H J = 6.6 Hz), 2.94 (q, 2 H J = 5.3 Hz), 2.71 (d, 6 H J = 5.8 Hz), 2.32 (m, 4 H), 1.83 (m, 4 H); MALDI-TOF-MS 978.7 (979.1 calcd for M + H).

ImImPy-γ-PyPyPy-β-Dp (1). Polyamide was prepared by machine-assisted solid phase synthesis protocols,[7] and 900 mg of resin was cleaved and purified to provide 1 as a white powder (69 mg, 48% recovery): UV λ_{max} 246 (43 300), 308 (54 200); ^{1}H NMR (DMSO-d_6)

(11) Kent, S. B. H. *Annu. Rev. Biochem.* **1988**, *57*, 957.

(12) Resin substitution can be calculated as L_{new}(mmol/g) = L_{old}/(1 + L_{old}(W_{new} − W_{old}) × 10^{-3}), where L is the loading (mmol of amine/g of resin), and W is the weight (gmol^{-1}) of the growing polyamide attached to the resin. See: Barlos, K.; Chatzi, O.; Gatos, D.; Stravropoulos, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 513.

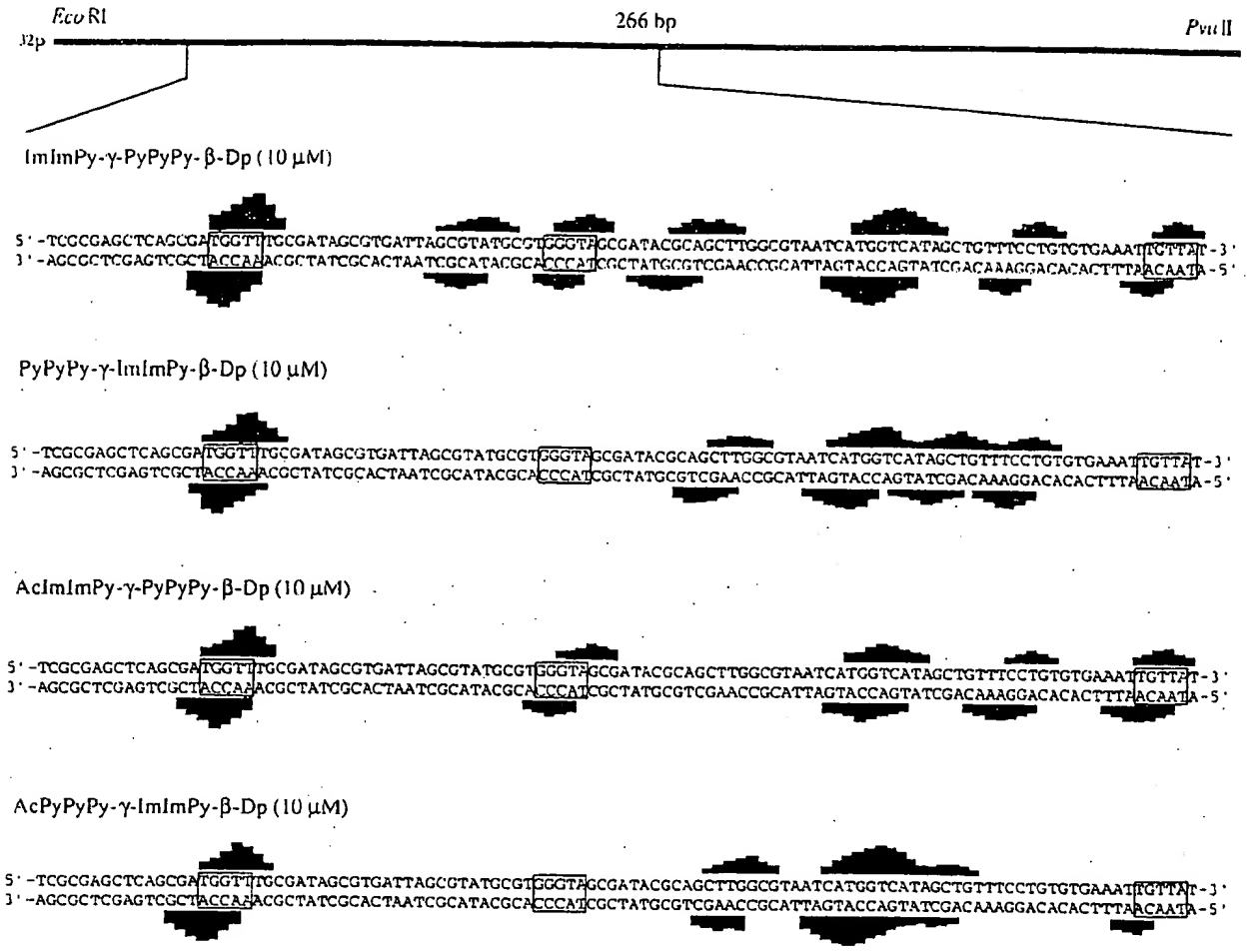


Figure 5. Histograms of cleavage protection (footprinting) data. (Top) Illustration of the 266 bp restriction fragment with the position of the sequence indicated. MPE·Fe^{II} protection patterns for polyamides at 10 μM concentration. Bar heights are proportional to the relative protection from cleavage at each band. Boxes represent equilibrium binding sites determined by the published model.[9] Only sites that were quantitated by DNase I footprint titrations are boxed.

Table 1. Equilibrium Association Constants (M^{-1})[a,b]

polyamide	match site 5′-aTGGTTt-3′	single mismatch sites 5′-tTGTTAt-3′	5′-tGGGTAg-3′
ImImPy-γ-PyPyPy-β-Dp	1.0×10^8 (0.1)	1.7×10^6 (0.6)	$\leq 1 \times 10^6$
PyPyPy-γ-ImImPy-β-Dp	1.6×10^7 (0.2)	$< 1 \times 10^5$	$< 1 \times 10^5$
AcImImPy-γ-PyPyPy-β-Dp	1.3×10^7 (0.7)	1.6×10^6 (1.1)	1.3×10^6 (0.8)
AcPyPyPy-γ-ImImPy-β-Dp	2.0×10^7 (0.3)	1.3×10^6 (0.5)	$\leq 1 \times 10^6$

[a] Values reported are the mean values measured from at least three footprint titration experiments, with the standard deviation for each data set indicated in parentheses. [b] The assays were performed at 22 °C at pH 7.0 in the presence of 10 mM Tris·HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

δ 10.31 (s, 1 H), 9.91 (s, 1 H), 9.90 (s, 1 H), 9.85 (s, 1 H), 9.75 (s, 1 H), 9.34 (br s, 1 H), 8.03 (m, 3 H), 7.56 (s, 1 H), 7.46 (s, 1 H), 7.21 (m, 2 H), 7.15 (m, 2 H), 7.07 (d, 1 H J = 1.2 Hz), 7.03 (d, 1 H, J = 1.3 Hz), 6.98 (d, 1 H, J = 1.2 Hz), 6.87 (m, 2 H), 4.02 (m, 6 H), 3.96 (m, 6 H), 3.87 (m, 6 H), 3.75 (q, 2 H, J = 4.9 Hz), 3.36 (q, 2 H, J = 4.0 Hz), 3.20 (q, 2 H, J = 4.7 Hz), 3.01 (q, 2 H, J = 5.1 Hz), 2.71 (d, 6 H, J = 4.8 Hz), 2.42 (m, 4 H), 1.80 (m, 4 H) MALDI-TOF-MS 978.8 (979.1 calcd for M + H).

AcImImPy-γ-PyPyPy-β-Dp (3). Polyamide was prepared by manual solid phase protocols and isolated as a white powder (8 mg, 28% recovery): UV λ_{max} 246 (43 400), 312 (50 200); ^{1}H NMR (DMSO-d_6) δ 10.35 (s, 1 H), 10.30 (s, 1 H), 9.97 (s, 1 H), 9.90 (s, 1 H), 9.82 (s, 1 H), 9.30 (s, 1 H), 9.2 (br s, 1H), 8.02 (m, 3 H), 7.52 (s, 1 H), 7.48 (s, 1 H), 7.21 (m, 2 H), 7.16 (d, 1 H, J = 1.1 Hz), 7.11 (d, 1 H, J = 1.2 Hz), 7.04 (d, 1 H, J = 1.1 Hz), 6.97 (d, 1 H, J = 1.3 Hz), 6.92 (d, 1 H, J = 1.4 Hz), 6.87 (d, 1 H, J = 1.2 Hz), 3.99 (s, 3 H), 3.97 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.80 (s, 3 H), 3.79 (s, 3 H), 3.47 (q, 2 H, J = 4.7 Hz), 3.30 (q, 2 H, J = 4.6 Hz), 3.20 (q, 2 H, J = 5.0 Hz), 3.05 (q, 2 H, J = 5.1 Hz), 2.75 (d, 6 H, J = 4.1 Hz), 2.27 (m, 4 H), 2.03 (s, 3 H), 1.74 (m, 4 H); MALDI-TOF-MS 1036.4 (1036.1 calcd for M + H).

AcPyPyPy-γ-ImImPy-β-Dp (4). Polyamide was prepared by machine-assisted solid phase protocols[7] as a white powder (14 mg, 48% recovery): UV λ_{max} 246 (44 400), 312 (52 300); ^{1}H NMR (DMSO-d_6) 10.32 (s, 1 H), 10.28 (s, 1 H), 9.89 (m, 2 H), 9.82 (s, 1 H), 9.18 (s, 1 H), 9.10 (br s, 1 H), 8.03 (m, 3 H), 7.55 (s, 1 H), 7.52 (s, 1 H), 7.21 (d, 1 H, J = 1.1 Hz), 7.18 (d, 1 H, J = 7.16 Hz), 7.15 (d, 1 H, J = 1.0 Hz), 7.12 (d, 1 H, J = 1.0 Hz), 7.02 (d, 1 H, J = 1.0 Hz), 6.92 (d, 1 H, J = 1.1 Hz), 6.87 (d, 1 H, J = 1.1 Hz), 6.84 (d, 1 H, J = 1.0 Hz), 3.97 (s, 3 H), 3.93 (s, 3 H), 3.87 (s, 3 H), 3.80 (s, 3 H), 3.78 (m, 6 H), 3.35 (q, 2 H, J = 5.6 Hz), 3.19 (q, 2 H, J = 5.3 Hz), 3.08 (q, 2 H, J = 5.7 Hz), 2.87 (q, 2 H, J = 5.8 Hz), 2.71 (d, 6 H, J = 4.0 Hz), 2.33 (m, 4 H), 1.99 (s, 3 H), 1.74 (m, 4 H); MALDI-TOF-MS 1036.2 (1036.1 calcd for M + H).

Construction of Plasmid DNA. Using T4 DNA ligase, the plasmid pMEPGG was constructed by ligation of an insert, 5′-GATCGC-

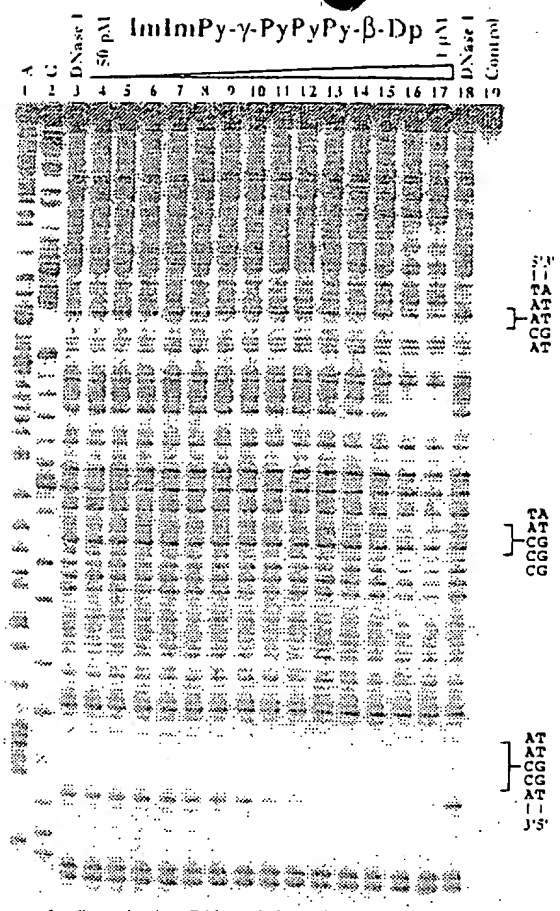


Figure 6. Quantitative DNase I footprint titration experiment with ImImPy-γ-PyPyPy-β-Dp (1) on the 3′-^{32}P-labeled 266 base pair *Eco*RI/*Pvu*II restriction fragment from plasmid pMEPGG. Lane 1: A reaction. Lane 2: G reaction. Lanes 3 and 18: DNase I standard. Lanes 4−17: 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM ImImPy-γ-PyPyPy-β-Dp (1), respectively. Lane 19: intact DNA. The 5′-TGGTT-3′, 5′-GGGTA-3′, and 5′-TGTTA-3′ sites which were analyzed are shown on the right side of the autoradiogram. All reactions contain 10 kpm restriction fragment, 10 mM Tris·HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

GAGCTCAGCGATGGTTTGCGATAGCGTGATTAGC-GTATGCGTGGGTAGCGATACGC-3′ and 5′-GCGTATCGCTAC-CCACGCATACGCTAAT-CACGCTATCGCAAACCATCGCTGA-GCTCGCGATC-3′, into pUC 19 previously cleaved with *Bam*HI and *Hind*III. Ligation products were used to transform Epicurian Coli XL 1 Blue competent cells (Stratagene). Colonies were selected for α-complementation on 25 mL Luria-Bertani medium agar plates containing 50 μg/mL ampicillin and treated with XGAL and IPTG solutions. Large-scale plasmid purification was performed with Qiagen purification kits. Plasmid DNA concentration was determined at 260 nm using the relation 1 OD unit = 50 μg/mL duplex DNA. The plasmid was linearized with *Eco*RI, followed by treatment with either Klenow, deoxyadenosine 5′-[α-^{32}P]triphosphate (Amersham), and thymidine 5′-[α-^{32}P]triphosphate for 3′ labeling or calf alkaline phosphatase and subsequent 5′ end labeling with T4 polynucleotide kinase and γ-[^{32}P]dATP. The 3′- or 5′-end-labeled fragment was then digested with *Pvu*II and isolated by nondenaturing gel electrophoresis. The 3′- or 5′-^{32}P-end-labeled 266 base pair *Eco*RI/*Pvu*II restriction fragment was used in all experiments described here. Chemical sequencing reactions were performed according to published protocols.[13] Standard protocols were used for all DNA manipulations.[14]

Identification of Binding Sites by MPE·Fe^{II} Footprinting. All reactions were carried out in a total volume of 40 μL with final concentrations of species as indicated in parentheses. The ligands were added to solutions of radiolabeled restriction fragment (10 000 cpm), calf thymus DNA (100 μM bp), Tris-acetate (25 mM, pH 7.0), and NaCl (10 mM) and incubated for 1 h at 22 °C. A 50 μM MPE·Fe^{II} solution was prepared by mixing 100 μL of a 100 μM MPE solution with a freshly prepared 100 μM ferrous ammonium sulfate solution. Footprinting reactions were initiated by the addition of MPE·Fe^{II} (5 μM), followed 5 min later by the addition of dithiothreitol (5 mM), and allowed to proceed for 15 min at 22 °C. Reactions were stopped by ethanol precipitation, resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer, and electrophoresed on 8% polyacrylamide denaturing gels (5% cross-link, 7 M urea) at 2000 V for 1 h. The gels were analyzed using storage phosphor technology.

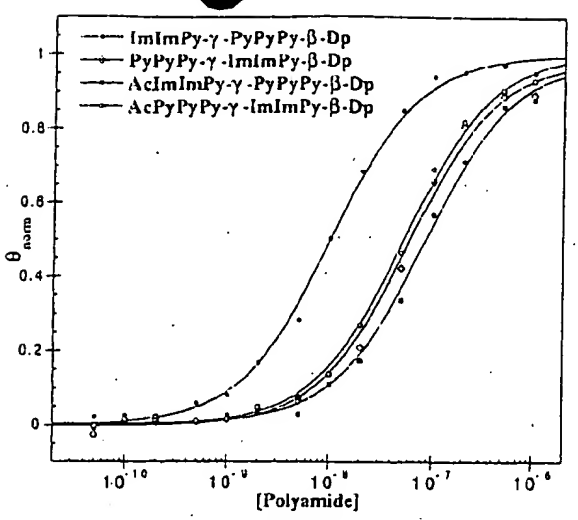


Figure 7. Data for the quantitative DNase I footprint titration experiments for the four polyamides 1−4 in complex with the designated 5′-TGGTT-3′ site. The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography and processed as described in the Experimental Section. The data points for ImImPy-γ-PyPyPy-β-Dp (1), PyPyPy-γ-ImImPy-β-Dp (2), AcImImPy-γ-PyPyPy-β-Dp (3), and AcPyPyPy-γ-ImImPy-β-Dp (4) are indicated by ●, ◇, ■, and ○, respectively. The solid curves are the best-fit Langmuir binding titration isotherms obtained from a nonlinear least squares algorithm using eq 2.

Analysis of Energetics by Quantitative DNase I Footprint Titration. All reactions were executed in a total volume of 40 μL with final concentrations of each species as indicated. The ligands, ranging from 50 pM to 1 μM, were added to solutions of radiolabeled restriction fragment (10 000 cpm), Tris·HCl (10 mM, pH 7.0), KCl (10 mM), $MgCl_2$ (10 mM), and $CaCl_2$ (5 mM) and incubated for 4 h at 22 °C. Footprinting reactions were initiated by the addition of 4 μL of a stock solution of DNase I (0.025 unit/mL) containing 1 mM dithiothreitol and allowed to proceed for 6 min at 22 °C. The reaction mixtures were stopped by addition of a 3 M sodium acetate solution containing 50 mM EDTA and ethanol precipitated. The reactions were resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer and electrophoresed on 8% polyacrylamide denaturing gels (5% cross-link, 7 M urea) at 2000 V for 1 h. The footprint titration gels were dried and quantitated using storage phosphor technology.

Equilibrium association constants were determined as previously described.[6,10] The data were analyzed by performing volume integrations of the 5′-TGGTT-3′, 5′-TGTTA-3′, and 5′-GGGTA-3′ sites and a reference site. Binding sites are assumed to be independent and noninteracting as they are separated by at least one full turn of the double helix. The apparent DNA target site saturation, θ_{app}, was calculated for each concentration of polyamide using the following equation:

(13) (a) Iverson, B. L.; Dervan, P. B. *Nucleic Acids Res.* 1987, *15*, 7823−7830. (b) Maxam, A. M.; Gilbert, W. S. *Methods Enzymol.* 1980, *65*, 499−560.

(14) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.

$$\theta_{app} = 1 - \frac{I_{tot}/I_{ref}}{I_{tot}^{o}/I_{ref}^{o}} \quad (1)$$

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}^{o} and I_{ref}^{o} correspond to those values for a DNase I control lane to which no polyamide has been added. The ($[L]_{tot}$, θ_{app}) data points were fitted to a Langmuir binding isotherm (eq 2, $n = 1$) by minimizing the difference between θ_{app} and θ_{fit}, using the modified Hill equation:

$$\theta_{fit} = \theta_{min} + (\theta_{max} - \theta_{min})\frac{K_a^{n}[L]_{tot}^{n}}{1 + K_a^{n}[L]_{tot}^{n}} \quad (2)$$

where $[L]_{tot}$ corresponds to the total polyamide concentration. K_a corresponds to the association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. The concentration of DNA used for quantitative footprint titrations is ≤50 pM, which justifies the assumption that free ligand concentration is approximately equal to total ligand concentration.[10] Data were fitted using a nonlinear least squares fitting procedure of KaleidaGraph software (version 2.1, Abelbeck software) running on a Power Macintosh 6100/60AV computer with K_a, θ_{max}, and θ_{min} as the adjustable parameters. The goodness-of-fit of the binding curve to the data points is evaluated by the correlation coefficient, with $R > 0.97$ as the criterion for an acceptable fit. At least three sets of acceptable data were used in determining each association constant. All lanes from each gel were used unless visual inspection revealed a data point to be obviously flawed relative to neighboring points. The data were normalized using the following equation:

$$\theta_{norm} = \frac{\theta_{app} - \theta_{min}}{\theta_{max} - \theta_{min}} \quad (3)$$

Quantitation by Storage Phosphor Technology Autoradiography. Photostimulable storage phosphorimaging plates (Kodak Storage Phosphor Screen S0230 obtained from Molecular Dynamics) were pressed flat against gel samples and exposed in the dark at 22 °C for 12−16 h. A Molecular Dynamics 400S PhosphorImager was used to obtain all data from the storage screens. The data were analyzed by performing volume integrations of all bands using the ImageQuant version 3.2 software running on an AST Premium 386/33 computer.

Acknowledgment. We are grateful to the National Institutes of Health (GM-27681) for research support, the National Institutes of Health for a research service award to M.E.P., and the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B.

JA9607289

Extension of Sequence-Specific Recognition in the Minor Groove of DNA by Pyrrole−Imidazole Polyamides to 9−13 Base Pairs

John W. Trauger, Eldon E. Baird, Milan Mrksich, and Peter B. Dervan*

Contribution from the Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125

Received March 6, 1996[⊗]

Abstract: The sequence-specific recognition of the minor groove of DNA by pyrrole−imidazole polyamides has been extended to 9−13 base pairs (bp). Four polyamides, ImPyPy-Py-PyPyPy-Dp, ImPyPy-G-PyPyPy-Dp, ImPyPy-β-PyPyPy-Dp, and ImPyPy-γ-PyPyPy-Dp (Im = *N*-methylimidazole, Py = *N*-methylpyrrole, Dp = *N,N*-dimethylaminopropylamide, G = glycine, β = β-alanine, and γ = γ-aminobutyric acid), were synthesized and characterized with respect to their DNA-binding affinities and specificities at sequences of composition 5′-(A,T)G-(A,T)$_5$C(A,T)-3′ (9 bp) and 5′-(A,T)$_5$G(A,T)C(A,T)$_5$-3′ (13 bp). In both sequence contexts, the β-alanine-linked compound ImPyPy-β-PyPyPy-Dp has the highest binding affinity of the four polyamides, binding the 9 bp site 5′-TGTTAAACA-3′ ($K_a = 8 \times 10^8$ M^{-1}) and the 13 bp site 5′-AAAAAGACAAAAA-3′ ($K_a = 5 \times 10^9$ M^{-1}) with affinities higher than the formally *N*-methylpyrrole-linked polyamide ImPyPy-Py-PyPyPy-Dp by factors of ~8 and ~85, respectively (10 mM Tris·HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0). The binding data for ImPyPy-γ-PyPyPy-Dp, which has been shown previously to bind DNA in a "hairpin" conformation, indicates that γ-aminobutyric acid does not effectively link polyamide subunits in an extended conformation. These results expand the binding site size targetable with pyrrole−imidazole polyamides and provide structural elements that will facilitate the design of new polyamides targeted to other DNA sequences.

Introduction

The development of 2:1 pyrrole−imidazole polyamide−DNA complexes provides a new model for the design of molecules for sequence-specific recognition in the minor groove of DNA. The polyamide ImPyPy-Dp was shown to specifically bind the mixed A,T/G,C sequence 5′-(A,T)G(A,T)C(A,T)-3′ as a side-by-side antiparallel dimer.[1] In this complex, each polyamide makes specific contacts with one strand on the floor of the minor groove such that the sequence specificity depends on the sequence of side-by-side amino acid pairings. A side-by-side pairing of imidazole opposite pyrrole recognizes G·C base pairs, while a side-by-side pairing of pyrrole opposite imidazole recognizes C·G base pairs.[1] A side-by-side pyrrole−pyrrole pairing is partially degenerate and targets both A·T and T·A base pairs.[1,2] The generality of the 2:1 model has been demonstrated by targeting other sequences of mixed A,T/G,C composition.[3−5] ImPyPy-Dp and distamycin (PyPyPy) bind simultaneously to a 5′-(A,T)G(A,T)$_2$-3′ site as an antiparallel heterodimer.[3] A PyImPy polyamide and distamycin bind 5′-(A,T)$_2$G(A,T)$_2$-3′[4] and ImPyImPy-Dp targets 5′-(A,T)GCGC-(A,T)-3′.[5]

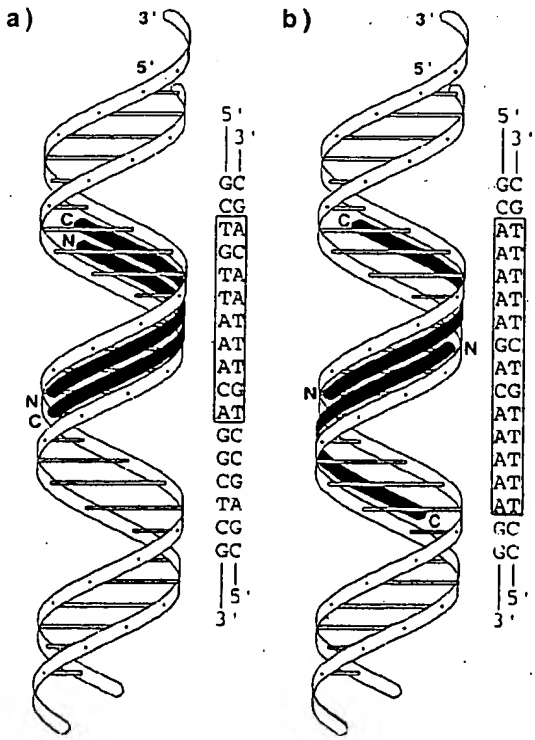


Figure 1. Ribbon models of (a) 9 bp "overlapped" and (b) 13 bp "slipped" 2:1 polyamide−DNA complexes.

The binding affinity and sequence specificity of a noncovalent antiparallel homodimeric or heterodimeric polyamide−DNA

[⊗] Abstract published in *Advance ACS Abstracts*, June 15, 1996.

(1) (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385.

(2) (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Nature Struct. Biol.* **1994**, *1*, 169.

(3) (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572−2576. (b) Geierstanger, B. H.; Jacobsen, J.-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994**, *33*, 3055.

(4) Geierstanger, B. H.; Dwyer, T. J.; Batlini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.

(5) (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646. (b) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 3325.

S0002-7863(96)00726-3 CCC: $12.00

4-9-97

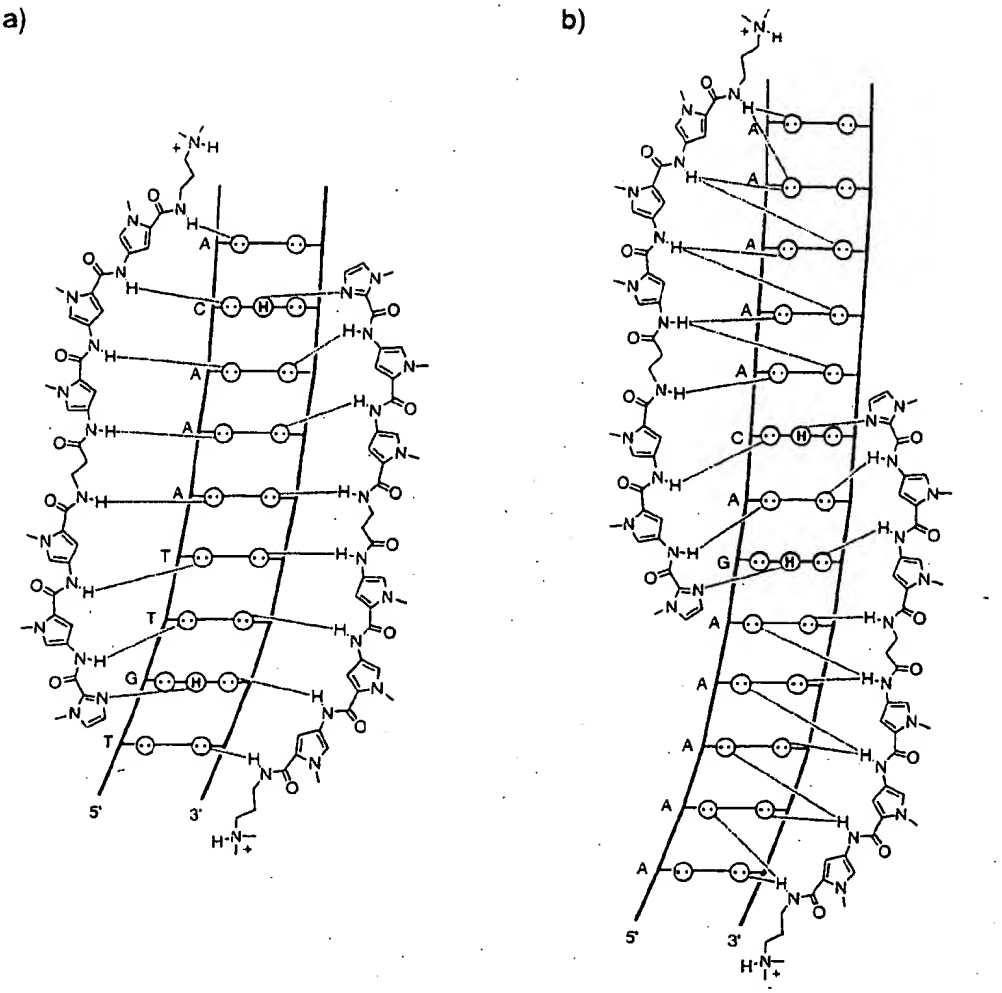


Figure 2. (Top) Complexes of ImPyPy-β-PyPyPy-Dp with the targeted sites (a) 5′-TGTTAAACA-3′ (9 bp "overlapped") and (b) 5′-AAAAAGACAAAAA-3′ (13 bp, "slipped"). Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dashed lines. (Bottom) Complexes of ImPyPy-X-PyPyPy-Dp, where X = Py, G, and β, with (a) 5′-TGTTAAACA-3′ and (b) 5′-AAAAAGACAAAAA-3′. The shaded and light circles represent imidazole and pyrrole rings, respectively, and the diamond represents the internal amino acid X. The specifically targeted guanines are highlighted.

complex can be increased by covalently linking the two polyamides.[6,7] The DNA-binding properties of the polyamides ImPyPy-G-PyPyPy-Dp, ImPyPy-β-PyPyPy-Dp, and ImPyPy-γ-PyPyPy-Dp, in which the terminal carboxyl group of ImPyPy and the terminal amine of PyPyPy-Dp are connected with glycine (G), β-alanine (β), and γ-aminobutyric acid (γ), respectively, were recently reported.[7,8] The γ-aminobutyric acid-linked polyamide bound the designated target site 5′-TGTTA-3′ with high affinity and sequence specificity and exhibited a binding isotherm in quantitative footprinting experiments consistent with formation of an intramolecular "hairpin" complex in which the polyamide folds back on itself.[7] Modeling suggested that the glycine- and β-alanine-linked polyamides do not favorably bind as "hairpins" in the minor groove of DNA. Moreover, these polyamides exhibited cooperative binding isotherms in quantitative footprinting experiments, consistent with two polyamides binding in extended conformations as *intermolecular dimers*.[7,8] It appears that the glycine- and β-alanine-linked polyamides disfavor binding in the hairpin conformation and prefer to bind in an extended conformation.[7,8] In a formal sense, there are multiple extended binding motifs (and hence multiple binding site sequences) for polyamides of sequence composition Im(Py)$_x$-Dp, as discussed below.

"Overlapped" and "Slipped" Binding Modes. We report here the DNA-binding affinities of four polyamides having the

(6) (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* 1993, *115*, 9892–9899. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* 1993, *115*, 9900. (c) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* 1994, *116*, 3663. (d) Singh, M. P.; Plouvier, B.; Hill, G. C.; Gueck, J.; Pon, R. T.; Lown, J. W. *J. Am. Chem. Soc.* 1994, *116*, 2006.

(7) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* 1994, *116*, 7983.

(8) Mrksich, M. Ph.D. Thesis, California Institute of Technology, 1994.

general sequence ImPyPy-X-PyPyPy-Dp, where X = Py, G, β, or γ, to the 9 bp site 5′-TGTTAAACA-3′ and to the 13 bp sites 5′-AAAAAGACAAAAA-3′ and 5′-ATATAGACATATA-3′. The polyamides having internal *N*-methylpyrrole, glycine, and β-alanine residues were anticipated to bind the 9 and 13 bp sites in an extended conformation. It was not clear at the outset if the γ-aminobutyric acid-linked polyamide ImPyPy-γ-PyPyPy-Dp would bind exclusively in an extended or "hairpin" conformation to the targeted sites. For ImPyPy-X-PyPyPy-Dp polyamides binding in an extended conformation, the polyamide−DNA complexes expected to form at the 9 bp and 13 bp target sites represent two distinct binding modes, which we refer to as "overlapped" and "slipped", respectively. In the "overlapped" (9 bp) binding mode, two ImPyPy-X-PyPyPy-Dp polyamides bind directly opposite one another (Figures 1a and 2a). The "slipped" (13 bp) binding mode integrates the 2:1 and 1:1 polyamide−DNA binding motifs at a single site. In this binding mode, the ImPyPy moieties of two ImPyPy-X-PyPyPy-Dp polyamides bind the central 5′-AGACA-3′ sequence in a 2:1 manner as in the ImPyPy homodimer,[1] and the PyPyPy moieties of the polyamides bind the all-A,T flanking sequences as in the 1:1 complexes of distamycin (Figures 1b and 2b). The structure of the complex formed by the polyamide ImPyPy-G-PyPyPy-Dp with a 13 bp target site has been characterized by 2D NMR.[9]

In the 9 bp "overlapped" and 13 bp "slipped" binding sites described above, the G·C and C·G base pairs are separated by one and five A,T base pairs, respectively. While we have concentrated here on these sites, we note that "partially slipped" sites of 10, 11, and 12 bp in which the G·C and C·G base pairs are separated by two, three, and four A,T base pairs, respectively, are also potential binding sites of the polyamides studied here.

Studies of the energetics of distamycin binding have shown that while the binding affinities are similar for complexation to poly[d(A-T)]·poly[d(A-T)] and poly[d(A)]·poly[d(T)], the origins of these binding affinities are different.[10] Binding to the alternating copolymer is enthalpy driven, while binding to the homopolymer is entropy driven.[10] However, not all 5 bp sites $(A,T)_5$ are bound with equal affinity. By quantitative footprinting experiments, the distamycin analog Ac-PyPyPy-Dp (Ac = acetyl) was shown to bind the sites 5′-AATAA-3′ and 5′-TTAAT-3′ with 2-fold and 14-fold lower affinity, respectively, relative to the site 5′-AAAAA-3′.[1c] On the basis of this result, we anticipated that the 13 bp site 5′-AAAAAGACAAAAA-3′ may be bound with higher affinity than the 13 bp site 5′-ATATAGACATATA-3′.

The binding affinities of polyamides **1**−**4** (Figure 3) for the three targeted sites 5′-TGTTAAACA-3′, 5′-AAAAAGACAAAAA-3′, and 5′-ATATAGACATATA-3′ were determined by quantitative DNase I footprint titration experiments.

Results

Synthesis of Polyamides. The polyamides ImPyPy-Py-PyPyPy-Dp (**1**),[11] ImPyPy-G-PyPyPy-Dp (**2**),[7] ImPyPy-β-Py-PyPy-Dp (**3**),[7] and ImPyPy-γ-PyPyPy-Dp (**4**)[7] were prepared as described previously.

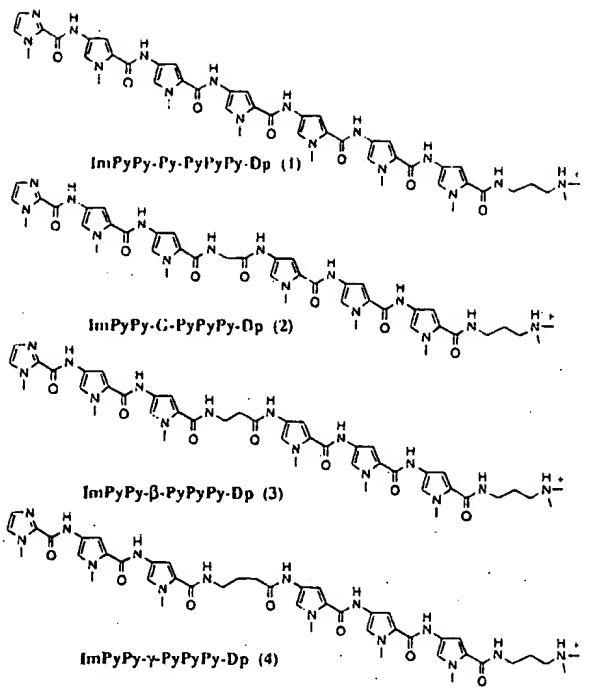


Figure 3. Structures of the four polyamides ImPyPy-X-PyPyPy-Dp, where X = *N*-methylpyrrole (Py), glycine (G), β-alanine (β), and γ-aminobutyric acid (γ).

Footprinting. Quantitative DNase I footprinting[12,13] on the 3′-^{32}P-labeled 281 bp pJT3 *Afl*II/*Fsp*I restriction fragment (Figures 4 and 5) (10 mM Tris·HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, 22 °C) reveals that, of the four polyamides ImPyPy-X-PyPyPy-Dp, three (X = Py, G, β) bind to both the 9 bp "overlapped" site 5′-TGTTAAACA-3′ and the 13 bp "slipped" site 5′-AAAAAGACAAAAA-3′ with equilibrium association constants (K_a) greater than $5 \times 10^7\ M^{-1}$ (Table 1) and display cooperative binding isotherms (eq 2, $n = 2$) at these sites *consistent with binding as intermolecular dimers* (Figure 6). The fact that the polyamides ImPyPy-G-PyPyPy-Dp and ImPyPy-β-PyPyPy-Dp bind in the 9 bp "overlapped" binding mode indicates that the internal glycine and β-alanine amino acids are accommodated opposite a second ligand in a 2:1 polyamide−DNA complex.

The polyamide ImPyPy-γ-PyPyPy-Dp binds the site 5′-TGTTAAACA-3′ with an equilibrium association constant K_a $1 \times 10^8\ M^{-1}$ and also binds the site 5′-AAAAAGACAAAAA-3′ with lower affinity ($K_a = 6 \times 10^6\ M^{-1}$). This compound displays binding isotherms (eq 2, $n = 1$) at these sites (Figure 6), *consistent with binding as an intramolecular hairpin* to the 5 bp "matched" sites 5′-TGTTA-3′ and 5′-AAACA-3′ (Figure 7) and to the 5 bp "single base pair mismatch" site 5′-AGACA-3′.[7] Significantly, it appears from these results that ImPyPy-γ-PyPyPy-Dp does not effectively link polyamide subunits in an extended conformation.

Binding Affinities. Comparison of the binding affinities of the four polyamides ImPyPy-X-PyPyPy-Dp, where X = Py, G, β, and γ, reveals that the internal amino acid X has a dramatic effect on complex stabilities (Table 1, Figure 6). The formally *N*-methylpyrrole-linked polyamide ImPyPy-Py-PyPyPy-Dp binds

(9) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Nature Struct. Biol.* **1996**, *3*, 321.

(10) Breslauer, K. J.; Remeta, D. P.; Chou, W.-Y.; Ferrante, R.; Curry, J.; Zaunczkowski, D.; Snyder, J. G.; Marky, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8922. (b) Marky, L. A.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4359. (c) Marky, L. A.; Kupke, K. J. *Biochemistry* **1989**, *28*, 9982.

(11) Kelly, J. J.; Baird, E. E.; Dervan, P. B., *Proc. Natl. Acad. Sci. U.S.A.*, in press.

(12) (a) Galas, D.; Schmitz, A. *Nucleic Acids Res.* **1978**, *5*, 3157. (b) Fox, K. R.; Waring, M. J. *Nucleic Acids Res.* **1984**, *12*, 9271.

(13) (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8462. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.

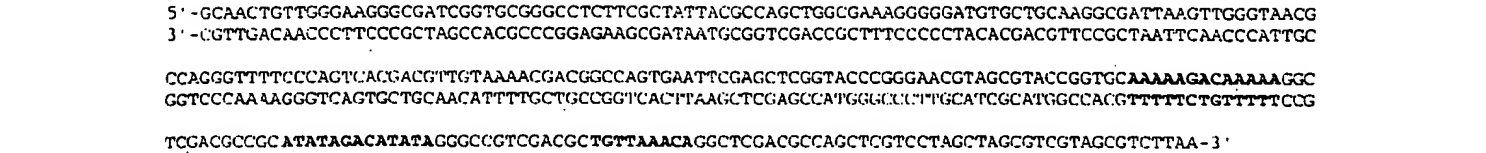


Figure 4. Sequence of the 281 bp pJT3 *Aflll/Espl* restriction fragment. The three binding sites that were analyzed in quantitative footprint titration experiments are indicated.

Table 1. Association Constants (M^{-1}) for Polyamides ImPyPy-X-PyPyPy-Dp, Where X = *N*-Methylpyrrole (Py), Glycine (G), β-Alanine (β), and γ-Aminobutyric Acid (γ)[a,b]

	polyamide			
binding site	ImPyPy-Py-PyPyPy-Dp	ImPyPy-G-PyPyPy-Dp	ImPyPy-β-PyPyPy-Dp	ImPyPy-γ-PyPyPy-Dp
5'-TGTTAAACA-3'	$9.7\ (\pm 2.3) \times 10^7$	$1.4\ (\pm 0.1) \times 10^8$	$7.8\ (\pm 0.6) \times 10^8$	$1.4\ (\pm 0.3) \times 10^8$
5'-AAAAAGACAAAAA-3'	$5.4\ (\pm 1.5) \times 10^7$	$1.1\ (\pm 0.1) \times 10^8$	$\geq 4.7\ (\pm 0.7) \times 10^9$	$6.4\ (\pm 0.6) \times 10^6$
5'-ATATAGACATATA-3'	$3.6\ (\pm 0.5) \times 10^7$	$6.6\ (\pm 0.4) \times 10^6$	$1.0\ (\pm 0.1) \times 10^9$	$4.6\ (\pm 0.5) \times 10^6$

[a] The reported association constants are the mean values obtained from three DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses. [b] The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris·HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

5'-TGTTAAACA-3' and 5'-AAAAAGACAAAAA-3' with equilibrium association constants $K_a = 1 \times 10^8\ M^{-1}$ and $5 \times 10^7\ M^{-1}$, respectively. The glycine-linked polyamide ImPyPy-G-PyPyPy-Dp binds both 5'-TGTTAAACA-3' and 5'-AAAAAGACAAAAA-3' with affinities similar (equal and 2-fold higher, respectively) to those of ImPyPy-Py-PyPyPy-Dp. In contrast, the β-alanine linked polyamide ImPyPy-β-PyPyPy-Dp binds 5'-TGTTAAACA-3' and 5'-AAAAAGACAAAAA-3' with affinities higher than those of ImPyPy-Py-PyPyPy-Dp by factors of approximately 8 ($K_a = 8 \times 10^8\ M^{-1}$) and 85 ($K_a = 5 \times 10^9\ M^{-1}$), respectively. Relative to ImPyPy-Py-PyPyPy-Dp, the hairpin-forming polyamide ImPyPy-γ-PyPyPy-Dp binds 5'-TGTTAAACA-3' (a matched hairpin binding site) and 5'-AAAAAGACAAAAA-3' (a mismatched hairpin binding site) with equal and 8-fold-lower affinities, respectively.

Specificity for 5'-AAAAAGACAAAAA-3' versus 5'-ATATAGACATATA-3'. Comparison of the binding affinities of the four polyamides at the 13 bp sites 5'-AAAAAGACAAAAA-3' and 5'-ATATAGACATATA-3' indicates that the specificity between the sites depends on the internal amino acid (Table 1). ImPyPy-G-PyPyPy-Dp and ImPyPy-β-PyPyPy-Dp are approximately 20-fold and ≥ 5-fold specific respectively for 5'-AAAAAGACAAAAA-3' versus 5'-ATATAGACATATA-3'. The polyamides ImPyPy-Py-PyPyPy-Dp and ImPyPy-γ-PyPyPy-Dp bind 5'-AAAAAGACAAAAA-3' and 5'-ATATAGACATATA-3' with similar affinities.

Discussion

Implications for the Design of Minor Groove Binding Molecules. The results presented here reveal that β-alanine is an optimal linker for joining two three-ring subunits in an extended conformation, providing a useful structural motif for the design of new polyamides targeted to sites longer than 8 bp. Recently, it has been shown that as the length of a polyamide having the general sequence Im(Py)$_x$-Dp increases beyond five rings (corresponding to a 7 bp binding site), the binding affinity ceases to increase with increasing polyamide length, indicating that the *N*-methylimidazole and *N*-methylpyrrole residues fail to maintain an appropriate base-pair register across the entire length of the polyamide–DNA complex.[11] The higher binding affinities observed here for ImPyPy-β-PyPyPy-Dp relative to ImPyPy-Py-PyPyPy-Dp indicate that the flexible β-alanine linker relieves the register mismatch, allowing both three-ring subunits to bind optimally. Notably, higher binding affinities are observed for ImPyPy-β-PyPyPy-Dp versus ImPyPy-Py-PyPyPy-Dp despite the higher conformational entropy and lower aromatic surface area of the β-alanine-linked polyamide. The observation here that β-alanine effectively links polyamide subunits within 2:1 polyamide–DNA complexes is consistent with the previously reported finding that β-alanine effectively links polyamide subunits within 1:1 polyamide–DNA complexes.[14]

From the standpoint of binding specificity, the observation here that a single compound can bind in multiple binding modes is problematic. The next generation of pyrrole–imidazole polyamides targeted to binding sites greater than 8 bp in length should incorporate constraints that specify a single binding mode.

The previously described γ-aminobutyric acid-based "hairpin" motif[7] complements the β-alanine-based "extended" motif described here. For ImPyPy-γ-PyPyPy-Dp, the binding isotherms and affinities observed here are consistent with the previous report that γ-aminobutyric acid effectively links polyamide subunits in a "hairpin" conformation[7] and indicate that γ-aminobutyric acid does not effectively link polyamide subunits in an extended conformation. Significantly, these observations indicate that (1) extended binding modes will not compromise the sequence specificity of hairpin-forming, γ-aminobutyric acid-linked polyamides, and (2) β-alanine and γ-aminobutyric acid linkers could be used within a single polyamide with predictable results.

The results reported here expand the binding site size targetable with pyrrole–imidazole polyamides and provide structural motifs that will facilitate the design of new pyrrole–imidazole polyamides targeted to other sequences.

Experimental Section

Materials. Restriction endonucleases were purchased from either New England Biolabs or Boeringer-Mannheim and used according to the manufacturer's protocol. *Escherichia coli* XL-1 Blue competent cells were obtained from Stratagene. Sequenase (version 2.0) was obtained from United States Biochemical, and DNase I was obtained from Pharmacia. [α-^{32}P]Thymidine 5'-triphosphate (≥ 3000 Ci/mmol) and [α-^{32}P]deoxyadenosine 5'-triphosphate (≥ 6000 Ci/mmol) were purchased from DuPont NEN.

Construction of Plasmid DNA. Plasmid pJT3 was prepared by standard methods. Briefly, plasmid pJT1 was prepared by hybridization of two complementary sets of synthetic oligonucleotides: 5'-CCGG-

(14) (a) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 7564. (b) Griffin, J. H.; Dervan, P. B. Unpublished results.

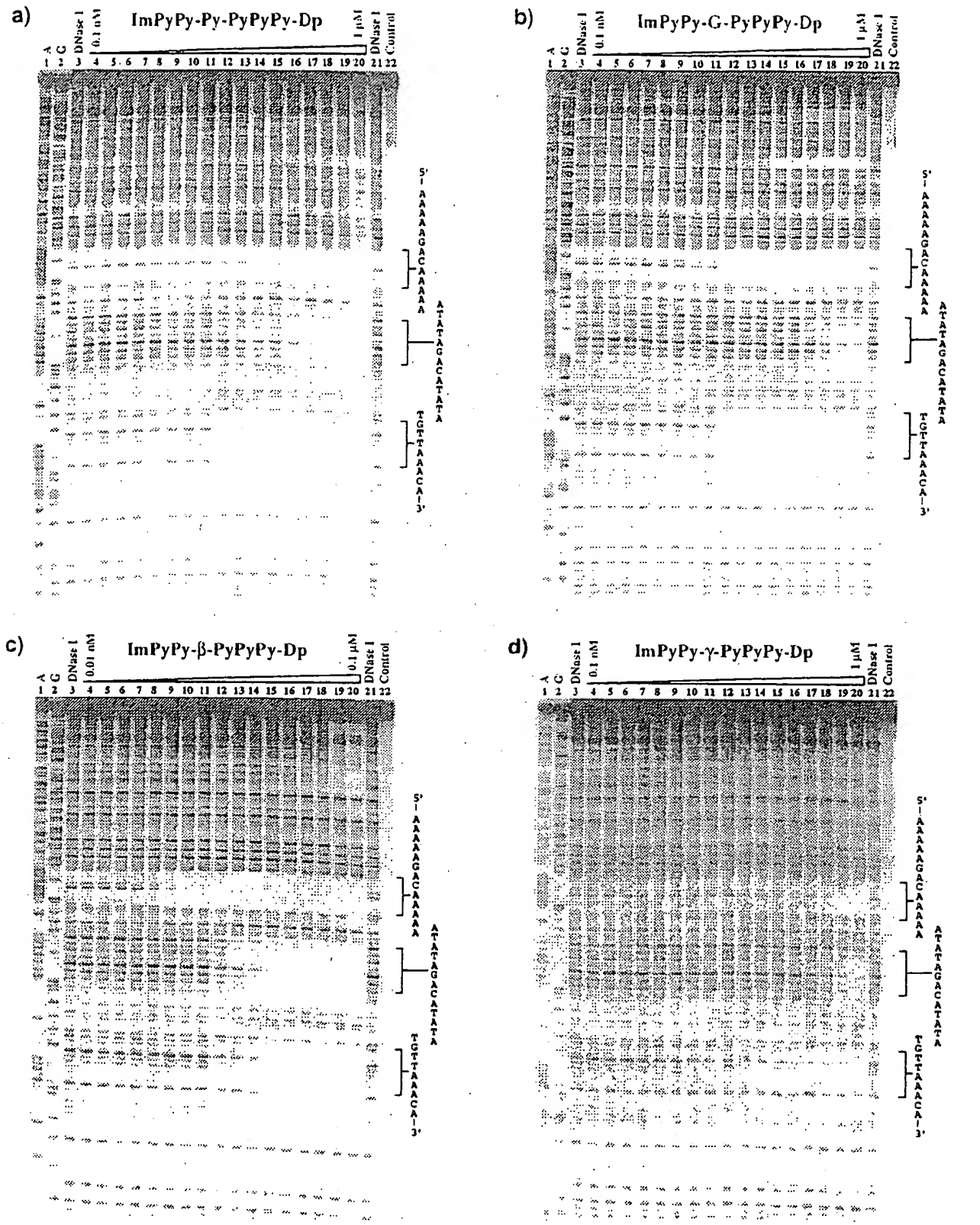


Figure 5. Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated by DNase I digestion in quantitative footprint titration experiments. Lanes 1 and 2, A and G sequencing lanes. Lanes 3 and 21: DNase I digestion products obtained in the absence of polyamide. Lanes 4–20: DNase I digestion products obtained in the presence of 0.1 nM (0.01 nM), 0.2 nM (0.02 nM), 0.5 nM (0.05 nM), 1 nM (0.1 nM), 1.5 nM (0.15 nM), 2.5 nM (0.25 nM), 4 nM (0.4 nM), 6.5 nM (0.65 nM), 10 nM (1 nM), 15 nM (1.5 nM), 25 nM (2.5 nM), 40 nM (4 nM), 65 nM (6.5 nM), 100 nM (10 nM), 200 nM (20 nM), 500 nM (50 nM), 1 μM (0.1 μM) (concentrations used for polyamide ImPyPy-β-PyPyPy-Dp only are in parentheses). Lane 22: intact DNA. The targeted binding sites are indicated on the right side of the autoradiograms. All reactions contain 15 kcpm restriction fragment, 10 mM Tris·HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

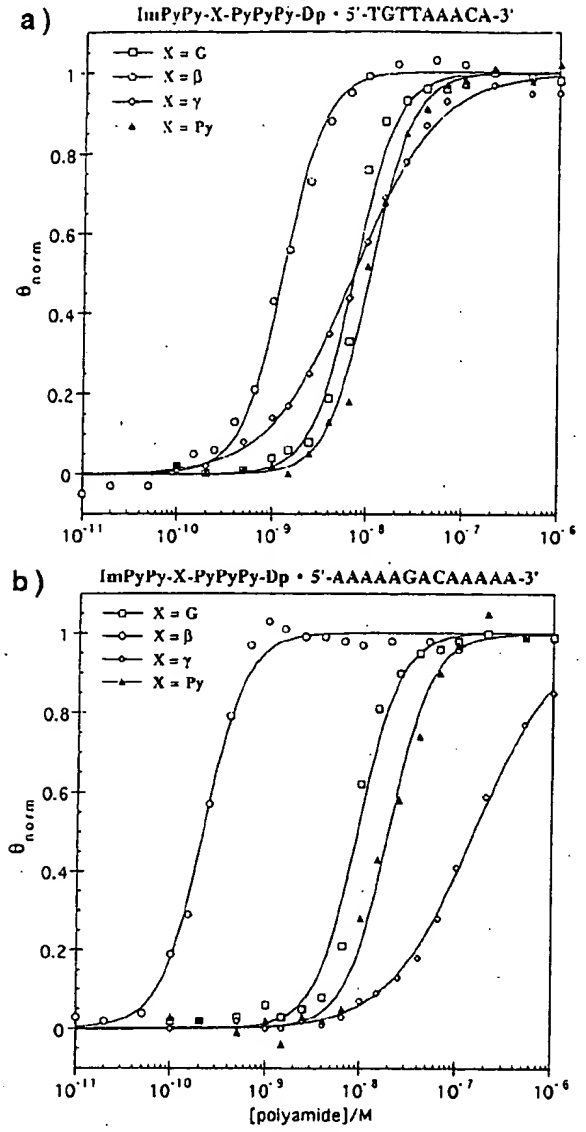


Figure 6. Data obtained from quantitative DNase I footprint titration experiments showing the effect of the internal amino acid X on binding of the polyamides ImPyPy-X-PyPyPy-Dp to the sites (a) 5′-TGTTAAACA-3′ and (b) 5′-AAAAAGACAAAAA-3′, where X = G (□), β (●), γ (◇), and Py (▲). The (θ_{norm}, $[L]_{tot}$) data points were obtained as described in the Experimental Section. Each data point is the average value obtained from three quantitative footprint titration experiments.

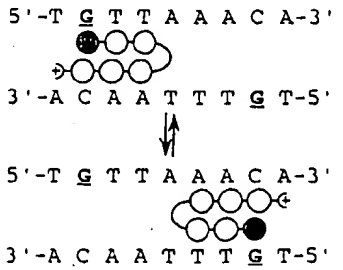


Figure 7. Model for the complex formed by ImPyPy-γ-PyPyPy-Dp with 5′-TGTTAAACA-3′ (5 bp, "hairpin").

GAACGTAGCGTACCGGTGCAAAAAGACAAAAAGG-CTCGA-3′ and 5′-GGCGTCGAGCCTTTTTGTCTTTTTGCACC-GGTACGCTACGTTC-3′, and 5′-CGCCGCATATAGACATATAG-GGCCCAGCTCGTCCTAGCTAGCGTCGTAGCGTCTTAAG-3′ and 5′-TCGACTTAAGACGCTACGACGCTAGCTAGGACGAGCT-GGGCCCTATATGTCTATATGC-3′. The resulting oligonucleotide duplexes were phosphorylated with ATP and T4 polynucleotide kinase and ligated to the large pUC19 *AvaI*/*Sal*I restriction fragment using T4 DNA ligase. *E. coli* XL-1 Blue competent cells were then transformed with the ligated plasmid, and plasmid DNA from ampicillin-resistant white colonies isolated using a Promega Maxi-Prep kit. Plasmid pJT3 was prepared in a similar manner, except that the following synthetic oligonucleotides were hybridized and cloned into the large *Apa*I/*Sal*I fragment of pJT1: 5′-GTCGACGCTGTTAAA-CAGGCTCGACGCCAGCTCGTCCTAGCTAGCGT-CGTAGCGTCTTAAGAG-3′ and 5′-TCGACTCTTAAGACGCTAC-GACGCTAGCTAGGACGAGCTGGCGTCGAGCCT-GTTTAACAGCGTCGACGGCC-3′. The presence of the desired insert was determined by restriction analysis and dideoxy sequencing. Plasmid DNA concentration was determined at 260 nm using the relation 1 OD unit = 50 μg/mL duplex DNA.

Preparation of ^{32}P-End-Labeled Restriction Fragments. Plasmid pJT3 was digested with *Afl*II, labeled at the 3′-end using Sequenase (version 2.0), and digested with *Fsp*I. The 281 bp restriction fragment was isolated by nondenaturing gel electrophoresis and used in all quantitative footprinting experiments described here. Chemical sequencing reactions were performed as described.[15,16] Standard techniques were employed for DNA manipulations.[17]

Quantitative DNase I Footprint Titration.[12,13] All reactions were executed in a total volume of 40 μL. We note explicitly that no carrier DNA was used in these reactions. A polyamide stock solution or H_2O (for reference lanes) was added to an assay buffer containing radiolabeled restriction fragment (15 000 cpm), affording final solution conditions of 10 mM Tris·HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, and either (i) 0.1 nM to 1 μM polyamide, for all polyamides except ImPyPy-β-PyPyPy-Dp, (ii) 0.01 nM to 0.1 μM polyamide for ImPyPy-β-PyPyPy-Dp, or (iii) no polyamide (for reference lanes). The solutions were allowed to equilibrate for 5 h at 22 °C. Footprinting reactions were initiated by the addition of 4 μL of a DNase I stock solution (at the appropriate concentration to give ~55% intact DNA) containing 1 mM dithiothreitol and allowed to proceed for 7 min at 22 °C. The reactions were stopped by the addition of 10 μL of a solution containing 1.25 M NaCl, 100 mM EDTA, and 0.2 mg/mL glycogen, and ethanol precipitated. The reaction mixtures were resuspended in 1X TBE/80% formamide loading buffer, denatured by heating at 85 °C for 10 min, and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% cross-link, 7 M urea) in 1X TBE at 2000 V. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics).

Quantitation and Data Analysis. Data from the footprint titration gels were obtained using a Molecular Dynamics 400S PhosphorImager followed by quantitation using ImageQuant software (Molecular Dynamics). Background-corrected volume integration of rectangles encompassing the footprint sites and a reference site at which DNase I reactivity was invariant across the titration generated values for the site intensities (I_{site}) and the reference intensity (I_{ref}). The apparent fractional occupancies (θ_{app}) of the sites were calculated using the equation

$$\theta_{app} = 1 - \frac{I_{site}/I_{ref}}{I_{site}^{\circ}/I_{ref}^{\circ}} \qquad (1)$$

where I°_{site} and I°_{ref} are the site and reference intensities, respectively, from a control lane to which no polyamide was added.

The ($[L]_{tot}$, θ_{app}) data points were fitted to a general Hill equation (eq 2) by minimizing the difference between θ_{app} and θ_{fit}:

$$\theta_{fit} = \theta_{min} + (\theta_{max} - \theta_{min})\frac{K_a''[L]_{tot}''}{1 + K_a''[L]_{tot}''} \qquad (2)$$

where $[L]_{tot}$ is the total polyamide concentration, K_a is the apparent association constant, and θ_{min} and θ_{max} are the experimentally deter-

(15) Iverson, B. L.; Dervan, P. B. *Nucleic Acids Res.* **1987**, *15*, 7823–7830.
(16) Maxam, A. M.; Gilbert, W. S. *Methods Enzymol.* **1980**, *65*, 499–560.
(17) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.

mined site saturation values when the site is unoccupied or saturated, respectively. The data were fitted using a nonlinear least squares fitting procedure with K_a, θ_{max}, and θ_{min} as the adjustable parameters, and with either $n = 2$ or $n = 1$ depending on which value of n gave the better fit. We note explicitly that treatment of the data in this manner does not represent an attempt to model a binding mechanism. Rather, we have chosen to compare values of the association constant, a parameter that represents the concentration of polyamide at which the binding site is half-saturated. The binding isotherms were normalized using the following equation:

$$\theta_{norm} = \frac{\theta_{app} - \theta_{min}}{\theta_{max} - \theta_{min}} \quad (3)$$

Three sets of data were used in determining each association constant.

The method for determining association constants used here involves the assumption that $[L]_{tot} \approx [L]_{free}$, where $[L]_{free}$ is the concentration of polyamide free in solution (unbound). For very high association constants this assumption becomes invalid, resulting in underestimated association constants. In the experiments described here, the DNA concentration is estimated to be ~50 pM. As a consequence, measured association constants of 1×10^{11} M^{-1} and 5×10^{10} M^{-1} underestimate the true association constants by factors of approximately ≤ 1.5 and 1.5–2, respectively.

Acknowledgment. We are grateful to the National Institutes of Health (Grant GM-27681) for research support, to the National Science Foundation for a predoctoral fellowship to J.W.T., and to the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B.

JA9607260

Hin Recombinase Bound to DNA: The Origin of Specificity in Major and Minor Groove Interactions

Jin-An Feng, Reid C. Johnson, Richard E. Dickerson*

The structure of the 52–amino acid DNA-binding domain of the prokaryotic Hin recombinase, complexed with a DNA recombination half-site, has been solved by x-ray crystallography at 2.3 angstrom resolution. The Hin domain consists of a three–α-helix bundle, with the carboxyl-terminal helix inserted into the major groove of DNA, and two flanking extended polypeptide chains that contact bases in the minor groove. The overall structure displays features resembling both a prototypical bacterial helix-turn-helix and the eukaryotic homeodomain, and in many respects is an intermediate between these two DNA-binding motifs. In addition, a new structural motif is seen: the six–amino acid carboxyl-terminal peptide of the Hin domain runs along the minor groove at the edge of the recombination site, with the peptide backbone facing the floor of the groove and side chains extending away toward the exterior. The x-ray structure provides an almost complete explanation for DNA mutant binding studies in the Hin system and for DNA specificity observed in the Hin-related family of DNA invertases.

The Hin recombinase catalyzes a DNA inversion reaction in the *Salmonella* chromosome (*1*, *2*). This site-specific recombination reaction controls the alternate expression of two flagellin genes by reversibly switching the orientation of a promoter. During the process of inverting the 1-kb segment of DNA, Hin proteins bind to left and right recombination sites (*hixL* and *hixR*, respectively) located at the boundaries of the invertible DNA segment. The *hixL* and *hixR* sites with their bound Hin protein then form a synaptic complex with a third *cis*-acting site, a recombinational enhancer, which itself is bound by two dimers of the 98–amino acid Fis protein. Formation of this invertasome complex (*3*–*5*) aligns the two recombination sites correctly and activates the Hin protein to initiate the exchange of DNA strands, leading to inversion of the intervening DNA.

Hin belongs to a family of bacterial DNA invertases or recombinases that includes Gin from phage Mu, Cin from phage P1, and Pin from the *e*14 prophage of *Escherichia coli*. In addition to sharing 66 to 80 percent sequence identity between pairs of sequences, this family of proteins can substitute functionally for one another in each biological system (*1*). These DNA invertases most likely constitute an evolutionary family not unlike the c-type cytochromes. The availability of DNA sequence information from the binding sites of all four systems makes the present study of Hin-DNA binding especially informative in elucidating principles of protein-DNA recognition.

Hin binds to each *hixL* and *hixR* recombination site as a dimer. The final 52 amino acids of the two chains (Fig. 1A) are involved in binding to a 26-bp recombination site (Fig. 1B) built from two 12-bp imperfect inverted repeats separated by a 2-bp core region where DNA strand exchange occurs (*6*–*8*). The amino-terminal 138–amino acid "catalytic" domain is positioned in part over the core nucleotides. Although the monomeric 52–amino acid peptide by itself can bind to a recombination half-site with low-to-moderate affinity (dissociation constant $K_d \approx 10^{-7}$), cooperative interactions between the amino-terminal domains of two intact Hin molecules are required for high-affinity binding ($K_d \approx 10^{-9}$) (*8*–*10*).

A large body of footprinting, mutation, and chemical derivatization data has indicated features of Hin-DNA interaction which are distinctive to prokaryotic DNA-binding proteins (*8*–*13*). Specific binding requires both major groove interactions involving a helix-turn-helix (HTH) α-helix motif and minor groove interactions involving the sequence Gly^{139}-Arg^{140}-Pro^{141}-Arg^{142}. If residues 139 and 140 are deleted from the carboxyl-terminal domain, for example, then sequence-specific binding to DNA is abolished. As Fig. 1A shows, these

A Amino Acid Sequences of DNA-Binding Domains of Enteric Invertases

```
α-Helix                1---------1      2------2      3------3
  Hin    GRPRAITKH.EQEQISRLLEK.GHP.RQQLAIIF.GIG.VSTLYRYF.PASSIKKRMN
  Gin    GRPFKLTKA.EQEQAGRLLAQ.GIP.RKQVALIY.DVA.LSTLYKKH.PAKRAHIENDDRIN
  Cin    GRRFKYQEE.TQQQMRRLLEK.GIP.RKQVAIIY.DVA.VSTLYKKF.PASSFQS
  Pin    GRRPKLTPE.QQAQAGRLIAA.GTP.RQKVAIIY.DVG.VSTLYKRF.PAGDK
         |        |                |              |        |
         139      148              162            173      181
```

B *hixL* binding site for Hin protein:

```
 -13          -8    * * *     -1+1              +8 * *   +13
5'-T-T-C-T-T-G-A-A-A-A-C-C-A-A-G-G-T-T-T-T-T-G-A-T-A-A-3'
3'-A-A-G-A-A-C-T-T-T-T-G-G-T-T-C-C-A-A-A-A-A-C-T-A-T-T-5'
                                  * * *
```

C Synthetic *hixL* half-site for crystallographic

```
                2  3  4  5  6  7  8  9 10 11 12 13 14 15
Strand 1:    5'-T--G--T--T--T--T--T*-G--A--T--A--A--G--A
Strand 2:         C--A--A--A--A--A--C--T--A--T--T--C*-T--A-5'
                 29 28 27 26 25 24 23 22 21 20 19 18 17 16
```

Fig. 1. Amino acid and DNA base sequences in the Hin recombinase family (*1*, *34*). (A) Amino acid sequence of the DNA-binding domain (the 52 carboxyl-terminal residues) of the Hin protein and corresponding regions in Gin, Cin, and Pin. Residues in boldface are identical in all four sequences or at least in three of the four. α Helices 1 to 3 as located in our Hin structure analysis are marked above the Hin sequence. For crystallography, this Hin fragment was synthesized manually or on an ABI 430A synthesizer by the solid-phase method as described previously (*8*). (B) Base sequence of the left DNA inversion site, *hixL*. Numbering is to either direction from the center of the inverted repeat. Asterisks mark purine bases that are protected from methylation by the binding of Hin. (C) The 14-bp synthetic *hixL* half-site as cocrystallized with the Hin 52-mer. Strand 1 of the duplex is numbered 2 through 15 to match the right half of the entire *hixL* site in (B). Bases in strand 2 are numbered separately. For ease of reference, note that base *n* in strand 1 is paired with base (32 − *n*) in strand 2. Base pairs will be referenced by strand 1 numbers alone: that is, "base pair 9" is understood to signify base pair G9-C23. Phosphates always are numbered according to the base that follows: Phosphate P9 occurs between T9 and G9, whereas across the helix on the other strand, phosphate P24 lies between C23 and A24. Hence phosphate *n* on strand 1 lies opposite phosphate (33 − *n*) on strand 2. Asterisks mark bases that were iodinated for purposes of multiple isomorphous replacement phase analysis.

The authors are with the Molecular Biology Institute and Department of Biological Chemistry, University of California, Los Angeles, CA 90024.

*To whom correspondence should be addressed at the Molecular Biology Institute.

two residues also are invariant among all four of the DNA invertases.

By comparison with pairwise amino acid sequence identities of 66 to 80 percent in the entire protein sequences, the carboxyl-terminal peptides shown in Fig. 1A have somewhat fewer identities, 49 to 62 percent, but still are obviously homologous proteins. The Hin interactions resemble those in the binding of DNA by the homeodomain in eukaryotes. Indeed, as noted by Affolter *et al.* (*14*), the amino acid sequence of the Hin binding domain can be aligned with that of the homeodomain of the *Drosophila* engrailed protein with a 27 percent sequence identity, sufficient to suggest a class resemblance.

We report here the 2.3 Å resolution crystal structure analysis of the 52–amino acid carboxyl-terminal DNA-binding domain of Hin complexed with a 14-bp DNA oligomer containing a half *hixL* binding site (Fig. 1C). The Hin peptide forms a three–α-helix core with an extended chain at each end. The core interacts with the major groove of DNA, whereas the flanking amino- and carboxyl-terminal chains extend along two regions of the minor groove. The carboxyl-terminal eight-residue tail of the Hin peptide crosses the phosphodiester backbone and is inserted in the minor groove in a manner that has not heretofore been encountered in DNA-protein complexes. The crystal structure provides a virtually complete explanation of base specificity experiments in solution, including mutant studies.

Structure of the complex. The structure of the Hin-DNA complex was solved by multiple isomorphous replacement with the use of three iodine derivatives (Table 1). A typical section of the final ($2F_o - F_c$) map is shown in Fig. 2. One 52–amino acid Hin domain binds to each 14-bp DNA half-site (Fig. 3). DNA helices consisting of the 13 complete base pairs are stacked end-to-end in the crystal, as in figure 2 of (*15*). The unpaired base T2 on strand 1 (Fig. 1C) swings up to make a Hoogsteen-like interaction with base pair 3, G3·C29. At the other end, unpaired base A16 on strand 2 is not defined in the electron density map and presumably is disordered. The *a* axis of the crystal, the direction of stacking of two DNA helices, has a length at room temperature of 86.4 Å = 26 × 3.32 Å. The 12–base pair steps along the helix produce a total rotation of 407° (average 33.9° per step), so that the nonbonded interhelix junction between base pair 15 (A15·T17) of one helix and base pair 3 (G3·C29) of the next requires a reverse twist of 360° − 407° = −47°.

The DNA half-site is a standard B-DNA helix, with the usual local variation in helix parameters (*16*, *17*). The helix is relatively straight and not curved around the protein domain as in CAP, trp repressor, 434 repressor, and Met repressor (*18*–*22*), and has been proposed for the Fis-DNA complex (*23*, *24*). However, an unusually large amount of DNA surface area is contacted by Hin. Upon binding Hin peptide, the DNA half-site monomer loses 1816 Å^2 of static solvent-accessible surface area.

The DNA half-site contains a short run of five AT base pairs (numbers 4 through 8) that could be regarded as a segment of A-tract DNA. Three frequent characteristics of A-tract DNA are a straight and unbent helix axis, narrow minor groove, and large propeller twist, large enough for the formation of bifurcated hydrogen bonds within the major groove between adjacent base pairs (*25*–*27*). However, in the Hin-DNA complex the minor groove maintains a uniform width of approximately 6.5 to 8.5 Å (minimal P-P atom separation across the groove, less 5.8 Å for two phosphate group radii), rather than the 3.5 to 4.5 Å typical of most A-tracts. Propeller twist is large all along the Hin-DNA complex, averaging

Table 1. Summary of crystallographic analysis. Crystals of Hin domain/DNA complex were grown by vapor diffusion against 15 to 18 percent PEG1500 as described elsewhere (*15*). The structure was determined initially to 3.2 Å resolution by multiple isomorphous replacement (MIR). Heavy atom parameters were refined and MIR phases calculated with the program HEAVY (*44*). The initial MIR map generated after solvent flattening (*45*) revealed clear density for B-form DNA and for most of the protein backbone density. This map was improved further by refining heavy atom parameters against solvent-flattened phases (*46*). After two additional cycles of phasing, solvent flattening, and heavy atom parameter refinement, the final MIR map, with mean phasing figure of merit of 0.55 for data between 20.0 and 3.2 Å, was used to build a model of the complex. However, it still was difficult to fit the amino acid sequence into many regions of the map. Only after phases were extended and modified to 2.8 Å by the method of Zhang and Main (*47*) did the map show clear density for side chains of some "marker" residues. At that point, all residues could be fitted unambiguously with the exception of the final eight carboxyl-terminal amino acids. Conventional positional refinement then was carried out to 2.8 Å with X-PLOR (*48*, *49*). To refine the model further against a new 2.3 Å data set collected at −150°C, rigid-body refinements were carried out in successive steps to 2.5 Å. After positional refinement and simulated annealing, the ($2F_o - F_c$) map was of sufficient quality to allow the last eight residues to be built into the minor groove of the DNA, following a clear and continuous density. Electron density for residues Ser^{163} and Ser^{164}, however, remained poorly defined. Refinement was extended to 2.3 Å in four cycles of simulated annealing with X-PLOR prior to tightly restrained B-factor refinement. At the present stage of refinement, the agreement of the atomic model to crystallographic data is R = 0.228 for 8.0 to 2.3 Å resolution data. Coordinates have been deposited with the Brookhaven Protein Data Bank and are available for immediate distribution.

Parameter	Native	Native (−150°C)	IdU[6]	IdC[18]	IdU[6] + IdC[18]
		*Unit cell dimensions**			
a (Å)	86.4	84.9	85.9	86.0	86.6
b (Å)	84.7	81.4	82.6	82.6	83.5
c (Å)	47.4	44.0	47.0	46.4	47.2
		Data collection statistics			
Resolution (Å)	2.8	2.3	3.2	3.2	3.5
Measured reflections	11673	17839	10387	7578	7578
Unique reflections	4177	5645	2703	2501	2020
Reflections > 2σ(*F*)	2653	5346			
Completeness (%)	92.6	80.1	92.1	87.5	87.0
R_{sym}† (%)	5.7	5.6	7.5	7.8	8.4
Mean isomorphous difference‡ (%)			18.4	16.9	18.3
		Phasing statistics			
Resolution (Å)			20–3.2	20–3.2	20–3.5
Cullis *R* factor			0.59	0.58	0.54
Phasing power§			1.80	1.74	2.42
		Refinement			
Resolution (Å)		8.0–2.3			
R factor‖		0.228			
Reflections with $F > 2\sigma$		5346			
Total number of atoms		978			
Water molecules	16				
Rms deviations					
Bond lengths (Å)		0.024			
Bond angles (deg.)	3.97				

*Space group *C*222, with one Hin-DNA complex per asymmetric unit. †$R_{sym} = \Sigma(|I - \langle I\rangle|)/\Sigma(I)$, where I is the observed intensity and $\langle I\rangle$ is the averaged intensity obtained from multiple observations of symmetry-related reflections. ‡The mean isomorphous difference is $\Sigma||F_{ph}| - |F_p||/\Sigma|F_p|$, where $|F_p|$ and $|F_{ph}|$ are structure factor amplitudes of protein and heavy atom derivative of the protein, respectively. §Phasing power is the mean amplitude of heavy atom structure factors, F_h, divided by E, the root-mean-square lack-of-closure error. The mean figure of merit, 20.0 to 3.2 Å, is 0.55. ‖Crystallographic *R* factor = $\Sigma(|F_o - F_c|)/\Sigma(F_o)$ where F_o and F_c are the observed and calculated structure magnitudes, respectively.

−16°, but is not systematically larger in the A-A-A-A-A region than elsewhere. Hence the five AT base pairs do not constitute a classical A-tract structure, perhaps because of binding to Hin.

Overall protein folding. The 52–amino acid DNA-binding domain of Hin consists of a compact bundle of three α-helices, with extended amino-terminal arm and carboxyl-terminal tail (Fig. 3). α-Helix 1 (Glu^{148} to Lys^{158}) lies parallel to the axis of the DNA, α-helix 2 (Arg^{162} to Phe^{169}) is nearly antiparallel to helix 1 with an angle of −25° between helix axes, and α-helix 3 (Val^{173} to Phe^{180}) is inserted in the major DNA groove parallel to the base pairs (not to the floor of the groove itself). The HTH motif formed by helices 2 and 3 is similar to those found in other prokaryotic regulatory DNA-binding proteins. The Hin HTH region can be superimposed on equivalent regions of Fis (*23*, *24*) and λ repressor (*28*) with root-mean-square deviations in Cα atomic positions of only 0.61 and 0.76 Å, respectively.

All three α helices are amphipathic, with hydrophobic residues packed tightly against one another in a hydrophobic core (Fig. 3A). Ile^{152} and Leu^{156} of helix 1 interdigitate with Leu^{165} and Phe^{169} of helix 2. Val^{173}, Leu^{176}, and Phe^{180} of helix 3 also point into the hydrophobic core, which is delineated by the blue polypeptide backbone regions in Fig. 3A. At the bottom in that view, Ile^{144} on the amino-terminal arm closes the hydrophobic pocket. These hydrophobic interactions appear to be the main forces stabilizing the folding of the Hin protein. They also are strongly conserved among the other DNA invertases, Gin, Cin, and Pin (Fig. 1A), supporting the inference that all four of these proteins are folded in the same way. Hydrophobic interactions are supplemented by hydrogen bonds between side chains: Arg^{162} (invariant among the four invertases), at the beginning of helix 2, is hydrogen-bonded to main chain carbonyl oxygens of Phe^{180} (the final residue of helix 3) and Pro^{181}. Most charged side chains of the protein are either in contact with DNA or exposed to the solvent.

Major groove protein-DNA interactions. α-Helix 3 is the DNA recognition helix for the Hin protein; helices 1 and 2 are too far from the DNA to permit direct interactions. Only Gln^{163} at the amino terminus of helix 2 makes an indirect DNA contact through a hydrogen bond to residue Tyr^{177} (invertase invariant) in helix 3, which in turn contacts phosphate P19 (Fig. 3B). Five interactions between helix 3 and DNA backbone phosphates position the recognition helix properly, and two–amino acid side chains, Ser^{174} and Arg^{178}, make specific bonds to the edges of base pairs G9·C23 and A10·T22, as detailed below. It is significant that four of the eight amino acids in helix 3 are completely invariant among the four DNA invertases, and another three are semi-invariant, with the same residue in three sequences and a closely related one in the fourth.

The five nonspecific interactions with DNA backbone phosphates are depicted in Fig. 3B. The side chain of Tyr^{177} reaches up to phosphate P19 on one edge of the major groove, whereas Tyr^{179} on the other side of the α helix reaches down to phosphate P8 directly across the groove on the other wall. One of the terminal $-NH_2$ groups of the Arg^{178} side chain donates a hydrogen bond to the remaining oxygen of phosphate P8. The side chain of Thr^{175} and the main chain amide of Gly^{172} anchor helix 3 even further by donating hydrogen bonds to phosphate P9. In contrast to other HTH DNA-binding proteins, all of these nonspecific anchoring contacts to DNA phosphates are made by residues of helix 3; the "three-point contact" made by residues Gly^{172}, Thr^{175}, Tyr^{177}, and Tyr^{179} efficiently braces helix 3 against the opening of the major groove of DNA in a position to make specific recognition interactions.

Specific base sequence recognition uses only two Hin side chains, Ser^{174} and Arg^{178}, and in part involves the mediation of water molecules (Fig. 4) in a manner that has been proposed for the *trp* repressor (*29*). The side chain of Ser^{174} donates a hydrogen bond to the N-7 atom of A10. One turn of α helix away from this position, the terminal $-NH_2$ of Arg^{178} that is not involved with P8 donates a similar hydrogen bond to the N-7 of G9. The Arg^{178} ε-imino nitrogen donates another hydrogen bond to bound water molecule 1, which in turn donates a bond to the O-4 of T22, essentially "reading" the fact that this base pair is indeed A10·T22 and not a G·C pair. The other proton of this water molecule bonds to water molecule 2, which receives another hydrogen bond from the N-6 amine of A21 (recognizing this as an A·T pair and not G·C), and donates hydrogen bonds to N-7 of the same base and to the main chain carbonyl oxygen of Ser^{174}. Ser^{174} is invariant among all four DNA invertases. Arg^{178} is substituted only by Lys, and when this happens, two basic side chains always appear adjacent at positions 178 and 179 (Fig. 1A), meaning that some of the hydrogen bonds of the Hin structure could well be preserved. Both of the bound waters have the tetrahedral geometry expected for water molecules, donating two hydrogen bonds and accepting two others. The fourth bond to water 1, not shown explicitly in Fig. 4B, must be to another water molecule not well localized in the electron-density map.

All of these specific interactions are drawn in Fig. 4, A and B. Together they recognize base pairs 9, 10, and 11 of the half recombination site. Indeed, the binding of Hin is particularly sensitive to alterations of base pairs 9 and 10 (*13*). Dimethyl sulfate modification of the N-7 position of G9 inhibits Hin binding (*8*, *10*). Methylation at the N-6 position of A10 by the deoxyadenosine methylase of *Salmonella* decreases binding affinity (*13*). Hin binding

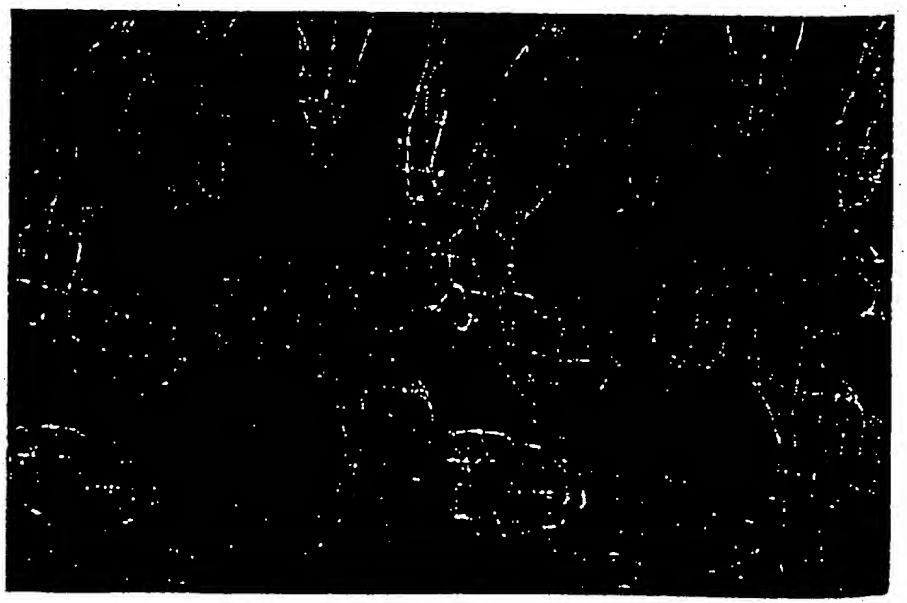


Fig. 2. Stereo view of the ($2F_o - F_c$) electron density map at 2.3 Å resolution (blue contours) showing portions of DNA base pairs 8 to 11 (top) and the region of helix 3 around Arg^{178} and Tyr^{179} (marked). Contour level is 1σ. The protein framework is in red, and DNA is in green.

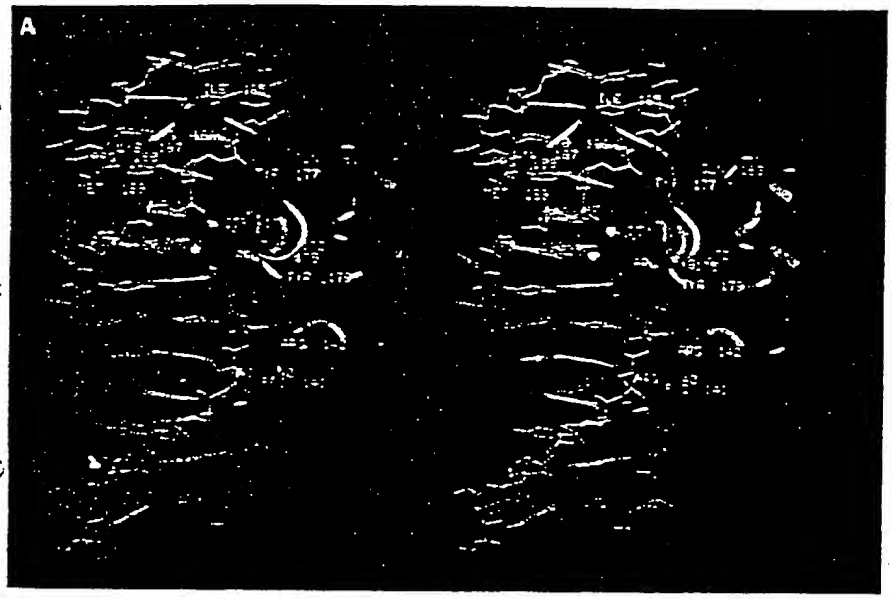
also is strongly and adversely affected by mutations at these sites, being inhibited by substitution of C at positions 9 and 11 or either C or T at position 10.

Minor groove protein-DNA interactions—the amino-terminal arm. Genetic and biochemical studies have demonstrated that contacts made by the amino-terminal arm of the Hin DNA-binding domain are at least as critical to DNA recognition as are those of helix 3 (*10*, *12*, *13*). Indeed, merely deleting Gly^{139} and Arg^{140} from the Hin DNA-binding peptide is sufficient to abolish specificity of binding to *hixL* (*12*). These residues are invariant in all of the DNA invertases.

The amino-terminal arm, Gly^{139}-His^{147}, adopts an extended conformation (Fig. 5). Clear electron density (Fig. 5A) allows Gly^{139} and Arg^{140} to be located unambiguously within the minor groove. The ε-imine of the Arg^{140} side chain donates a hydrogen bond to N-3 of A26. The unusually high 26° propeller twist of base pair 6 (T6·A26) permits a second hydrogen bond from the main chain amide of Arg^{140} to the O-2 of T6. If a G·C pair were to be substituted, this latter bond would become impossible, and the N-2 amine of guanine would push the Arg^{140} side chain away. Although the neighboring A·T base pairs are less propeller-twisted, the ability of an A·T pair to adopt such a large propeller contributes to the recognition process (*25*–*27*). Gly^{139} rests in close van der Waals contact with base pair 5; the main chain Cα atom of that residue is only 3.4 Å from the O-2 of T5 and 4.1 Å from the C-2 of A27. Introduction of an amine group at that locus, as in guanine, would push the Hin polypeptide chain up and away from the floor of the minor groove at that point. Each base pair substitution of A·T by G·C at positions 5 and 6 abolishes the binding affinity of Hin (*13*). Indeed, A·T base pairs at positions 5 and 6 are universally present in all of the recombination sites of various enteric inversion systems: *hixL* and *hixR*; *gixL* and *gixR*; *cixL* and *cixR*; and *pixL* and *pixR* (*1*) (Fig. 6). Biochemical footprinting experiments also show that both intact Hin and the Hin peptide protect adenines 5 and 6 from methylation (*10*).

Pro^{141} arches across one wall of the minor groove, and the ε-imino of Arg^{142} is hydrogen-bonded to phosphate P8, an interaction that may be important in directing the amino-terminal arm into the minor groove of the DNA. Hin, Gin, Cin, and Pin all have Pro at position 141 or 142, followed immediately by a basic Arg or Lys. A significant role may be played by Ile^{144}, which interacts with the hydrophobic core formed by packing helices 1 to 3 against one another. Ile^{144} may restrict movement of the amino-terminal arm, thus bringing Arg^{142} into proximity to phosphate P8. In other DNA invertases, position 144 is always a bulky hydrophobic side chain, either Leu or Tyr (Fig. 1A).

Minor groove protein-DNA interactions—the carboxyl-terminal tail. The carboxyl-terminal tail of the Hin polypeptide crosses the phosphodiester backbone at the



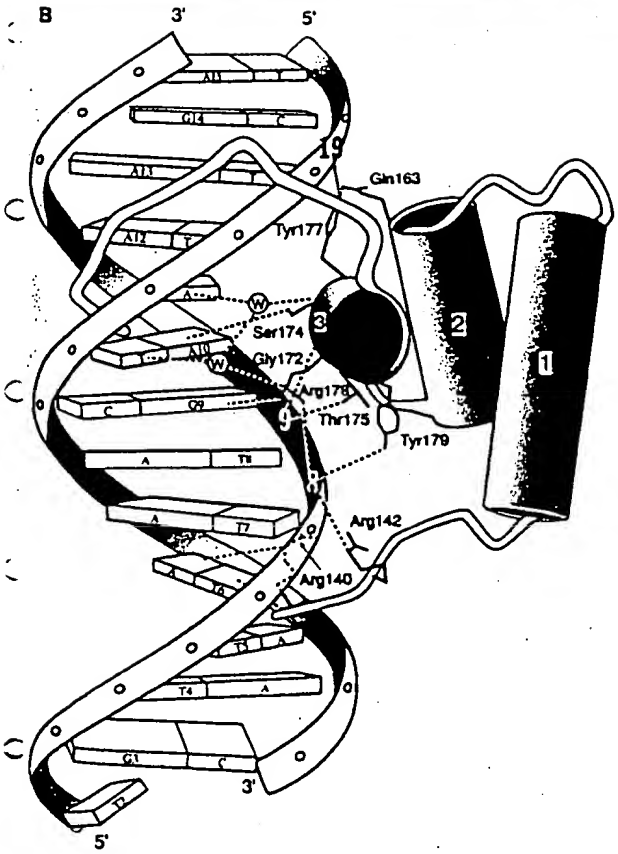


Fig. 3. (A) Stereo diagram of the Hin-DNA complex. DNA is in blue stick bonds. The path of the Hin polypeptide chain is shown as a flexible tubing: orange in general, but blue for hydrophobic residues. Side chains that contact the DNA are drawn in orange and labeled in yellow. Note the amino-terminal arm within the minor groove at lower right (Gly^{139}-Arg^{142}), the carboxyl-terminal tail in the minor groove at upper left (Ile^{185}-Asn^{190}), and helix 3 nested in the major groove. Pink spheres are two bound water molecules involved in protein-DNA contacts within the recognition site. Hydrogen bonds are in green. (B) Schematic drawing of the complex, in the same view as the stereo. Helices are numbered 1 to 3, and key side chains that interact with the DNA are identified. Open dots along backbone ribbons locate C1′ atoms. Phosphates 8, 9, and 19 are indicated specifically on the ribbons. Base pairs 9 to 11 are shown in stereo close-up in Fig. 4A, base pairs 4 to 8 in Fig. 5B, and pairs 9 to 15 in Fig. 7.

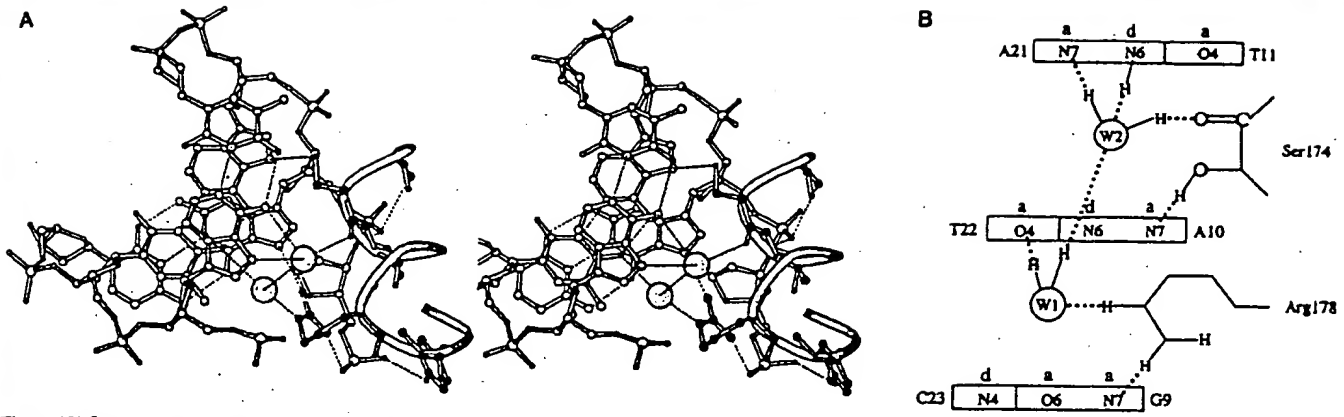


Fig. 4. (A) Stereo close-up of the interaction between DNA and helix 3 in the major groove, viewed approximately down the DNA helix axis. Base pair T11·A21 is nearest the viewer, with A10·T22 beneath, and G9·C23 farthest away. Strand 1 backbone continues to lower right through P9 and P8. Helix 3 is drawn as a smooth curve, with specific depiction, from top to bottom, of residues Gly^{172}, Ser^{174}, Thr^{175}, Arg^{178}, and Tyr^{179}. Two shaded spheres are bound water molecules. (B) Schematic of specific base pair recognition involving Ser^{174}, Arg^{178}, and two bound water molecules. Along base edges, "a" marks a hydrogen bond acceptor (ring N-7 or carbonyl O-4 or O-6) and "d" marks a hydrogen bond donor (N-4 or N-6 amine group).

outer edge of the recombination site and then follows the minor groove back toward the center of the 13-bp DNA helix (Figs. 3 and 7). The final six residues of the Hin polypeptide, Ile^{185}-Lys^{186}-Lys^{187}-Arg^{188}-Met^{189}-Asn^{190}, adopt an extended conformation and lie within the minor groove, but the side chains themselves make no contacts with the floor of the groove. Instead, they point outward, with the polypeptide backbone resting against base edges. At the point where the final six–amino acid residues dip into the minor groove, the main chain CO of Ile^{185} hydrogen bonds to the N-2 of G14. The main chain NH of Lys^{187} bonds to the O-2 of T20, and a little farther along, the main chain amide of Asn^{190} interacts with the O-2 of T22 while the side chain of the carboxyl-terminal residue bonds to the N-3 of A10. These interactions may be responsible for the large propeller twist and roll angles of base pairs 10 and 12 and the −16° bend of DNA toward the major groove. Consistent with the interactions just discussed, the N-3 of A10 is partially protected from dimethyl sulfate attack by Hin binding (*10*). In addition, Mack *et al.* (*11*) have noted that a Hin peptide lacking the last eight residues, when modified with EDTA·Fe, cleaved DNA with reduced efficiency as compared with a peptide containing the complete carboxyl terminus. This portion of the chain is variable among the DNA invertases: whereas Hin has six final amino acids, Gin has ten, Cin has three, and Pin has a lone Lys (Fig. 1A). This carboxyl-terminal tail is presumably supportive but not essential.

The hydrogen-bonded extension of the last six residues along the floor of the minor groove recalls AT-specific binding of minor groove drugs such as netropsin and distamycin (30, 31). Such binding involves an element of base specificity: if any of the base pairs 10 to 13 were G·C rather than A·T, then the tail of the Hin peptide would be pushed away from the floor of the groove. In another context, it has been proposed (32, 33) that an extended polypeptide containing repeats of SPKK (34) sequence may interact with DNA minor groove in a similar fashion to netropsin, with main chain amide nitrogen forming hydrogen bonds with base pairs in the minor groove. Our structure seems to provide a concrete example of such a model.

The molecular basis of specificity. Two aspects of the Hin system make it especially conducive to an understanding of the ef-

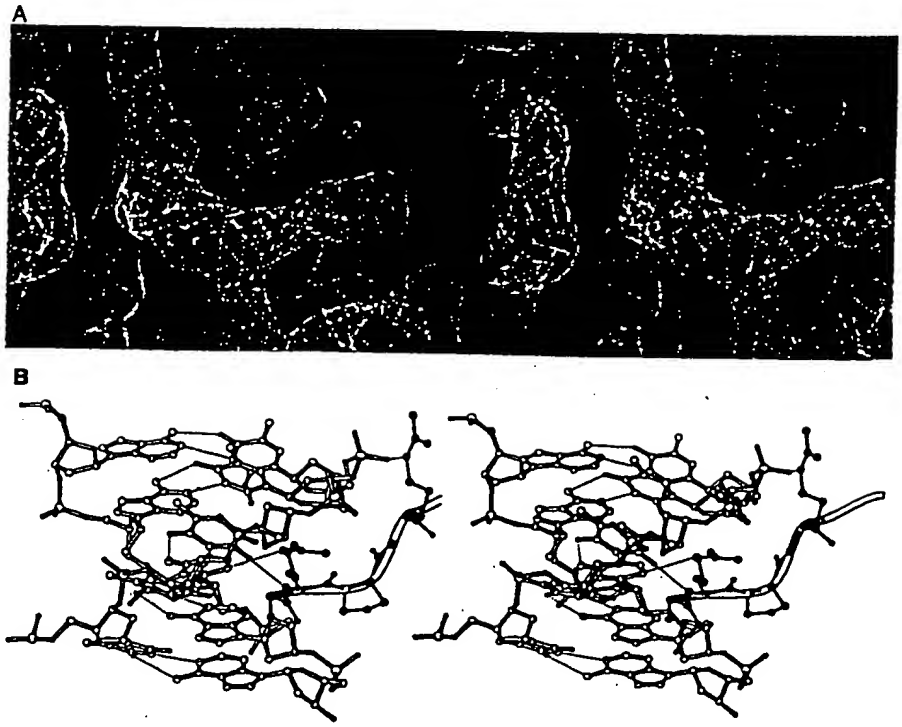


Fig. 5. Stereo views of the amino-terminal arm of the Hin peptide, in the minor groove of DNA. (A) The refined ($2F_o - F_c$) electron density map (blue framework) with minor groove vertical, showing Arg^{140}-Pro^{141}-Arg^{142} looping over the phosphate backbone toward the right. Protein is in red, and DNA is in green. (B) View along the minor groove, from the top in (A), showing the entire "A-tract" region, from T4·A28 at bottom through T8·A24 at top. This view is approximately that of the lower half of Fig. 3B. The ribbon extending from center toward upper right is the Hin peptide region Gly^{139}-Arg^{140}-Pro^{141}-Arg^{142}, with side chains drawn explicitly.

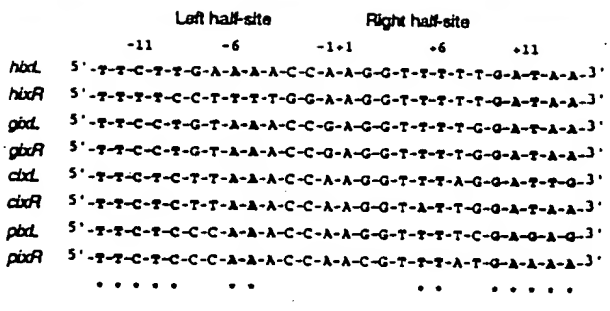


Fig. 6. Base sequences in enteric bacterial inversion sites: the left and right inversion sites from the Hin inversion system of *Salmonella*, the Gin inversion elements from bacteriophage Mu, Cin from phage P1, and Pin from the *e*14 prophage of *Escherichia coli* (*1*). Only one chain from each complex is shown; the other is complementary as in Fig. 1B. Each of the eight sites is built from two roughly symmetrical half-sites. Asterisks at bottom indicate positions that are especially important in site recognition by invertases.

fects of sequence on specificity. The first is the availability of binding data on 39 different base substitution mutations generated by Hughes *et al.* (*13*). They constructed a symmetrized *hix*C sequence in which the left half is given the complementary sequence to the right half shared by both *hixL* and *hixR* and established that this symmetrized *hix*C binds Hin fully as well as the wild-type *hixL* and *hixR*. (It is the *hix*C sequence that we used for crystallographic analysis.) They then constructed an exhaustive set of symmetrical mutants in the two halves of *hix*C, varying each of the 13 positions among all three of the other bases. Hence we have complete information about the strength of Hin binding with every possible single-base change in the optimal *hix*C sequence. The second favorable aspect is the existence of four homologous DNA inversion systems: Hin, Gin, Cin, and Pin. Taken together, these provide 8 complete recombination sites or 16 binding half-sites (Fig. 6). How far can our x-ray analysis of the Hin-DNA complex account for this wealth of data, and provide a molecular basis for DNA-protein specificity?

The Hin-DNA crystal structure shows that all three components of the Hin peptide, the amino-terminal arm, the HTH region, and the carboxyl-terminal tail, contribute to base sequence recognition. The phosphate backbone contacts of helix 3 help position Hin on the DNA, but one could easily imagine that Hin could slide along the DNA in a nonspecific manner until it encounters the correct local base sequence.

Interactions of Hin residues Ser^{174} and Arg^{178}, both direct and through intermediate water molecules (Fig. 4B) place restrictions on base pairs 9, 10, and 11. Some latitude in base sequence is possible if different arrangements of hydrogen bond donors and acceptors are permitted. Those rearrangements that are possible without losing the total number of hydrogen bonds are shown in Fig. 8. Base 9 is restricted to being a purine (G or A) by virtue of the hydrogen bond donated to ring atom N-7. In complete agreement with this model, of the 16 half-sites shown in Fig. 6 and listed in Table 2, G occurs 13 times at position 9 and A occurs 3 times. No pyrimidine is ever found at that locus. Replacement of G at position 9 by A or T (and C at position −9 by T or A) in the mutant studies of Hughes *et al.* (*13*) is acceptable, but C at position 9 reduces binding significantly. Our crystal structure shows why: G, A, or T at position 9 offer a hydrogen bond acceptor to Arg^{178}, whereas C offers a N-4 amine donor instead, and hence is disfavored. Even T is less favorable, because it positions the hydrogen bond acceptor differently and partially blocks it with its own C-5 methyl group.

Position 10 also must be a purine for the same reason as 9. The choice in 14 of the

Table 2. Frequency of occurrence of bases at key positions in bacterial inversion sites. Sequences are read in a 5′-to-3′ direction from the center of the inversion site as shown in Fig. 6. Hence, for left half-sites the other chain, not shown in Fig. 6, is tabulated. Allowed substitution data derive from the frequency of lysogenization in a P22 challenge phage assay at 100 μM isopropyl-β-d-thiogalactopyranoside (IPTG) concentration, table 1 of (*13*).

Position	Natural *hix*, *gix*, *cix*, and *pix* sites				Acceptable mutations in symmetrized *hixC* site (*13*)	DNA groove
	G	A	T	C		
5		2	14		Not G, not C	Minor
6		1	15		Not G, not C	Minor
9	13	3			Not C	Major
10	2	14			Not T, not C	Major
11	8	2	6		Not C	Major
12		15	1		Not C	Minor
13	2	14			All equivalent	Minor

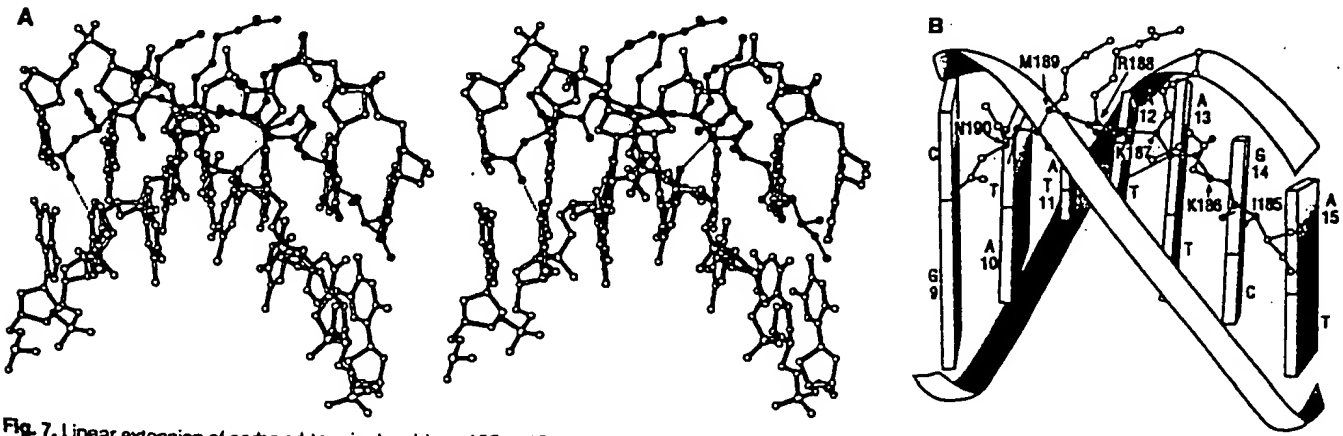


Fig. 7. Linear extension of carboxyl-terminal residues 185 to 190 down the minor groove. (**A**) Stereo pair representation. DNA base pair G9-C23 is at left, and A15-T17 at right. Amino acid side chains are identified in (**B**), which is a sketch from the same orientation. The main polypeptide chain atoms are in black dots; side chains are in open circles.

16 half-site sequences is A10, resulting in T22 at the other end of the base pair, with a hydrogen bond–accepting O-4 atom (Fig. 8A). However, G10-C22 is an acceptable minority choice in two sequences, and in that case the N-4 amine of C22 must donate a hydrogen bond to water 1 (Fig. 8B). Water 1 then would donate a hydrogen bond to another water molecule not shown here.

Base pair 11 is more variable than might have been expected, and for an interesting reason. Water molecule 2 in Fig. 8A accepts a hydrogen bond from the N-6 of A21 and donates a bond to N-7, but all that is required for a hydrogen balance is that this water molecule should donate one hydrogen and accept another. The two base atoms could just as easily be thymine O-4 and adenine N-6 as in Fig. 8A or cytosine N-4 and guanine O-6 as in Fig. 8B. The only combination not permitted would be guanine at position 21, with dual acceptors N-7 and O-6. For in that case, water 2 would not have enough protons to form the bond with the main chain carbonyl of Ser[174]. In other words, the requirement on the strand 1 side of base pair 11 is "not-C," and indeed this requirement is borne out in Table 2 both by the invertase family sequence comparisons and mutational substitutions.

The direction of the hydrogen bond between water molecules—water 1 donating to water 2—actually is firmly established. As Fig. 8C shows, if water 2 were to donate a bond both to water 1 and to the Ser[174] main chain carbonyl, then the two positions on base pair 11 would have to be adjacent donors, and no base pair shows this behavior. The full pattern of hydrogen bonds can be maintained only by arrangements as in Fig. 8, A and B.

A·T base pairs are favored at positions 12 and 13 because of a netropsin-like extension of the carboxyl-terminal six Hin residues down the floor of the minor groove (Fig. 7). Table 2 shows that A·T pairs indeed are overwhelmingly favored at these two loci, although mutant studies show position 13 to be more permissive. It could be that the similarity of sequences at this point among all of the enteric recombinases is a matter of evolutionary divergence from a common ancestral sequence, rather than convergence on function.

At the other end of the recognition domain, positions 4, 5, and 6 universally are A·T base pairs in all of the DNA inversion sites, and G·C pairs are strongly disfavored at positions 5 and 6 in the mutant studies. The x-ray structure shows the reason: Hin residues Gly[139] and Arg[140] are so intimately linked to the floor of the minor groove that there simply is no room for the C-2 amine group of guanine. As noted above, Gly[139] and Arg[140] are absolutely essential for sequence-specific binding of Hin to its DNA site.

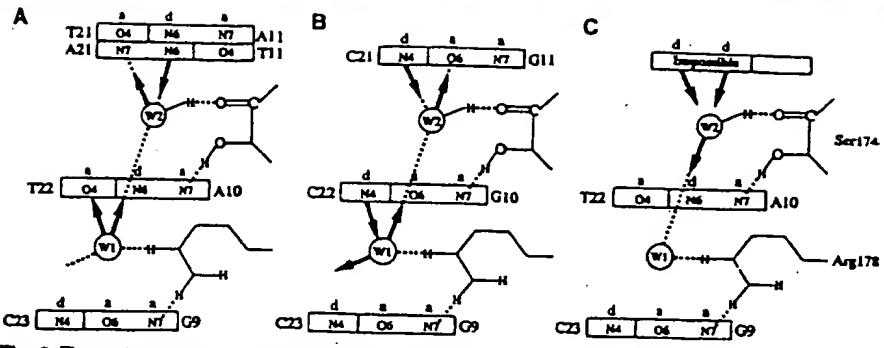


Fig. 8. The molecular basis for specificity, in the form of possible patterns of hydrogen bond donors and acceptors. (**A**) The pattern actually observed in the crystal structure. Base pair 11 could be reversed end-for-end without altering the pattern of hydrogen bond acceptor-donor-acceptor along its edge. (**B**) Hydrogen bond donors and acceptors could be reversed at base pair 11, allowing G11·C21. However, base 21 could not be G, with dual acceptors without breaking the bond between water 2 and the Ser[174] carbonyl. Hence C11·G21 is strongly disfavored. The acceptor at the central base pair can independently be replaced by a donor, permitting G10·C22 in place of A10·T22, but base 10 is still required to be a purine. (**C**) The hydrogen bond connecting the two water molecules cannot be reversed, for then, if the donation to Ser[174] carbonyl is maintained, base-pair 11 would have to contribute two adjacent donors, which is chemically impossible for any of the base pairs.

Fig. 9. Optimal base sequence for enteric inversion half-sites. Numbers mark the distance out from the center of the site as in Fig. 6. A/T = an A·T base pair in either orientation. N = any base.

5	6	7	8	9	10	11	12	13
(A/T)–	(A/T)–	N–	N–	(purine)–	A–	(not-C)–	(A/T)–	(weak A)

In summary, the recognition element of the enteric inversion half-sites appears to involve two A·T base pairs (5 and 6) recognized by amino acid residues Gly[139] and Arg[140] in the minor groove, two nonspecific base pairs (7 and 8), and then a five base sequence (9 to 13) recognized by helix 3 and the carboxyl-terminal tail in major and minor grooves, respectively. The optimal binding sequence is shown in Fig. 9. The Hin dimer has ~100-fold higher binding affinity to the full recombination site than does the Hin peptide binding to a half-site, and hence cooperative interactions by the Hin dimer may also contribute to recognition.

Hin-DNA versus other HTH DNA-binding complexes. The HTH motif occurs frequently in DNA-binding proteins, including prokaryotic regulators (*18, 20, 28, 35*), eukaryotic homeodomains (*36–38*), eukaryotic transcription factors such as HNF-3γ (*39*), the Oct-1 POU-specific domain, POUs (*40*), the third tandem repeat of the cMyb protein (*41*), and the globular domain of histone H5, GH5 (*42*). Complexes of these proteins with DNA show two principal variants that are represented in Fig. 10 by the λ repressor and the engrailed homeodomain. In all cases, recognition helix 3 fits into the major groove and helix 2 runs across the width of the groove. In homeodomain structures, helix 1 lies essentially antiparallel to helix 2, which also spans the width of the major groove. The residues preceding helix 1 are positioned to interact with the minor groove, and recognition helix 3 is oriented along the floor of the major groove. This pattern is persistent; the cluster of three helices is virtually superimposable from one homeodomain complex to the next. In contrast, prokaryotic regulatory proteins (*18, 20, 28, 35*) have similar dispositions of helices 2 and 3, but with the exception of *trp* repressor, helix 1 is swung out and away from the DNA (Fig. 10A). Helix 3, on the other hand, tends to lie parallel to the edges of base pairs in the prokaryotic regulators, rather than along the floor of the groove as in homeodomains.

The present Hin-DNA complex is intermediate between these two structures. Recognition helix 3 is parallel to base pair edges rather than to the groove itself, as with prokaryotic repressors, but helix 1 is nearly antiparallel to 2, with its amino terminus extending toward the minor groove where a nonhelical chain continuation contributes interactions essential to specific recognition. The alignment of helix 3 parallel to base pairs in Hin and repressors is functional: comparison of the structures of 434 repressor-operator, 434 Cro-operator, λ repressor-operator, and CAP-DNA complexes, shows that amino acids at positions *1*, *2*, and 6 of the helix make specific contacts with bases in the major groove in each case

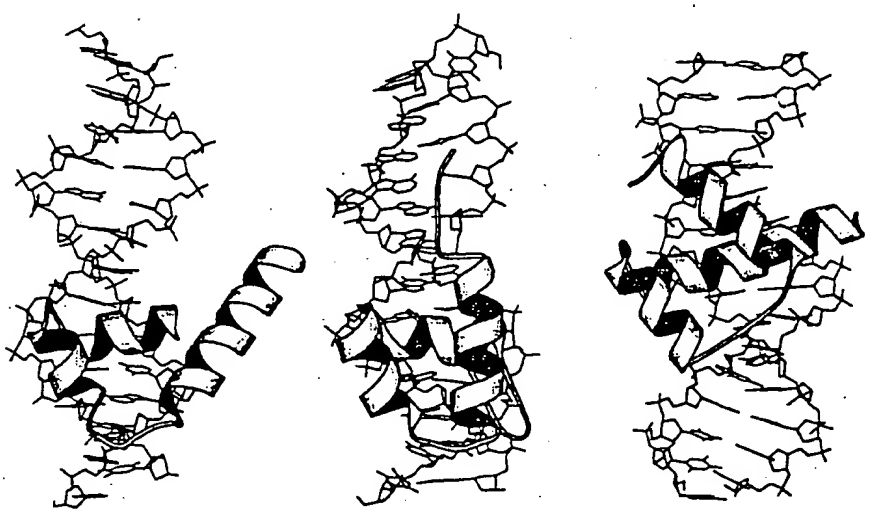


Fig. 10. Comparative interactions of a three-helix unit with the DNA double helix in the λ repressor–operator complex (**left**), the Hin-DNA complex (**center**), and the engrailed homeodomain complex (**right**). In each case, the carboxyl-terminal helix of the three is inserted into the major groove.

(43). Similarly, Hin uses positions 2 and 6 (Ser^{174} and Arg^{178}) for its primary recognition process.

The disposition of helix 1 with respect to the DNA is not completely identical in Hin and homeodomains; in the latter, helix 1 crosses perpendicular to the walls of the major groove, whereas Hin has helix 1 aligned parallel with the overall DNA axis. The two loops connecting helices are shorter in Hin than in the homeodomain proteins and more like those of prokaryotic repressors. In addition, the α helices themselves are shorter than their homeodomain equivalents, especially helix 3, which in both yeast α2 and *Drosophila* engrailed homeodomains are 17 amino acids long.

There also is a remarkable difference in minor groove binding by the sequence Arg-Pro-Arg in the amino-terminal arm of Hin and in the engrailed homeodomain. In Hin, Arg^{140} makes specific base contacts, whereas Arg^{142} anchors the two-pronged fork by binding to the phosphate backbone of DNA. In the engrailed homeodomain, Arg^3-Pro^4-Arg^5 inserts both Arg side chains into the minor groove and makes specific base contacts, and it is the adjacent Thr^6 that interacts with phosphate backbone. Thus, the same three–amino acid recognition module can interact with the minor groove in two profoundly different ways to generate contacts essential for protein-DNA binding.

In another significant difference, the amino-terminal arms of homeodomain proteins run along the minor groove in the direction of the HTH unit itself, whereas the amino-terminal arm of the Hin protein runs in an opposite direction (Figs. 3A and 9), toward what would be the center of the recombination site. Model-building of an intact recombinase bound to a complete *hix* site suggests that the positions cleaved by Hin during recombination are located on the opposite face of *hix* from the DNA-binding domains. It is likely that residues preceding the amino-terminal arm within an intact Hin protomer may continue running along the minor groove around the DNA and link with the catalytic domain that presumably is positioned over the cleavage site.

REFERENCES AND NOTES

1. A. C. Glasgow, K. T. Hughes, M. I. Simon, in *Mobile DNA*, D. E. Berg and M. M. Howe, Eds. (American Society for Microbiology, Washington, DC, 1989), pp. 637–659.
2. R. C. Johnson, *Curr. Opin. Genet. Dev.* 1, 404 (1991).
3. K. A. Heichman and R. C. Johnson, *Science* 249, 511 (1990).
4. K. A. Heichman, I. P. G. Moskowitz, R. C. Johnson, *Genes Dev.* 5, 1622 (1991).
5. I. P. G. Moskowitz, K. A. Heichman, R. C. Johnson, *ibid.*, p. 1635.
6. R. C. Johnson and M. I. Simon, *Cell* 41, 781 (1985).
7. R. C. Johnson and M. F. Bruist, *EMBO J.* 8, 1581 (1989).
8. M. F. Bruist, S. J. Horvath, L. E. Hood, T. A. Steitz, M. I. Simon, *Science* 235, 777 (1987).
9. J. P. Sluka, S. J. Horvath, M. F. Bruist, M. I. Simon, P. B. Dervan, *ibid.* 238, 1129 (1987).
10. A. C. Glasgow, M. F. Bruist, M. I. Simon, *J. Biol. Chem.* 264, 10072 (1989).
11. D. P. Mack *et al.*, *Biochem. J.* 29, 6561 (1990).
12. J. P. Sluka, S. J. Horvath, A. C. Glasgow, M. I. Simon, P. B. Dervan, *ibid.*, p. 6551.
13. K. T. Hughes, P. C. W. Gaines, J. E. Karlinsey, R. Vinayak, M. I. Simon, *EMBO J.* 11, 2695 (1992).
14. M. Affolter *et al.*, *Cell* 64, 879 (1991).
15. J.-A. Feng *et al.*, *J. Mol. Biol.* 232, 982 (1993).
16. G. G. Privé, K. Yanagi, R. E. Dickerson, *ibid.* 217, 177 (1991).
17. K. Grzeskowiak, K. Yanagi, G. G. Privé, R. E. Dickerson, *J. Biol. Chem.* 286, 8861 (1991).
18. S. C. Schultz, G. C. Shields, T. A. Steitz, *Science* 253, 1001 (1991).
19. Z. Otwinowski *et al.*, *Nature* 335, 321 (1988).
20. A. K. Aggarwal, D. W. Rodgers, M. Drottar, M. Ptashne, S. C. Harrison, *Science* 242, 899 (1988).
21. W. S. Somers and S. E. V. Phillips, *Nature* 359, 387 (1992).
22. T. A. Steitz, *Q. Rev. Biophys.* 23, 205 (1990).
23. H. S. Yuan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 88, 9558 (1991).
24. D. Kostrewa *et al.*, *Nature* 349, 178 (1991).
25. H. R. Drew *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 78, 2179 (1981).
26. H. C. M. Nelson, J. T. Finch, B. F. Luisi, A. Klug, *Nature* 330, 221 (1987).
27. M. Coll, C. A. Frederick, A. H.-J. Wang, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* 84, 8385 (1987).
28. S. R. Jordan and C. O. Pabo, *Science* 242, 893 (1988).
29. T. E. Haran, A. Joachimiak, P. B. Sigler, *EMBO J.* 11, 3021 (1992).
30. M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura, R. E. Dickerson, *Proc. Natl. Acad. Sci. U.S.A.* 82, 1376 (1985).
31. ——— *J. Mol. Biol.* 183, 553 (1985).
32. M. Suzuki, *EMBO J.* 8, 797 (1989).
33. M. E. Churchill and A. A. Travers, *Trends Biochem. Sci.* 16, 92 (1991).
34. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
35. A. Mondragon and S. C. Harrison, *J. Mol. Biol.* 219, 321 (1991).
36. C. R. Kissinger, B. S. Liu, E. Martin-Blanco, T. B. Kornberg, C. O. Pabo, *Cell* 63, 579 (1990).
37. C. Wolberger, A. K. Vershon, B. Liu, A. D. Johnson, C. O. Pabo, *ibid.* 67, 517 (1991).
38. G. Otting *et al.*, *EMBO J.* 9, 3085 (1990).
39. K. L. Clark, E. D. Halay, E. Lai, S. K. Burley, *Nature* 364, 412 (1993).
40. N. Dekker *et al.*, *ibid.* 372, 852 (1993); N. Assa-Munt, R. J. Mortishire-Smith, R. Aurora, W. Herr, P. E. Wright, *Cell* 73, 193 (1993).
41. K. Ogata, *Proc. Natl. Acad. Sci. U.S.A.* 89, 6428 (1992).
42. V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee, R. M. Sweet, *Nature* 362, 219 (1993).
43. S. C. Harrison and A. K. Aggarwal, *Annu. Rev. Biochem.* 59, 933 (1990).
44. T. C. Terwilliger, S.-H. Kim, D. Eisenberg, *Acta Crystallogr.* A43, 1 (1987).
45. B.-C. Wang, *Methods Enzymol.* 115, 90 (1985).
46. M. A. Rould, J. J. Perona, T. A. Steitz, *Acta Crystallogr.* A48, 751 (1992).
47. K. Y. J. Zhang and P. Main, *ibid.* A46, 377 (1991).
48. A. T. Brünger, J. Kuriyan, M. Karplus, *Science* 235, 458 (1987).
49. A. T. Brünger, *X-PLOR Manual*, version 3.1 (Yale Univ. Press, New Haven, CT, 1992).

50. We thank M. I. Simon and P. B. Dervan for providing Hin peptide, S. Finkel for help with DNA synthesis and purification, K. Zhang for many helpful discussions on phasing, and D. S. Goodsell for help with schematic drawings. Supported by NIH Program Project Grant GM-31299 (R.E.D.) and by GM-38509 (R.C.J.).

17 August 1993; accepted 15 December 1993

The Dictionary of

CELL BIOLOGY

Second Edition

EDITED BY

J. M. Lackie and J. A. T. Dow

Yamanouchi Research Institute, Littlemore Hospital, Oxford OX4 4XN, UK

and

Laboratory of Cell Biology, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

AUTHORS

S. E. Blackshaw[1]
C. T. Brett
A. S. G. Curtis
J. A. T. Dow
J. G. Edwards[2]
J. M. Lackie
A. J. Lawrence
G. R. Moores

[1] Second edition only
[2] First edition only

ACADEMIC PRESS

Harcourt Brace & Company, Publishers
London San Diego New York
Boston Sydney Tokyo Toronto

ACADEMIC PRESS LIMITED
24–28 Oval Road
London NW1 7DX

US Edition published by
ACADEMIC PRESS INC.
San Diego, CA 92101

First published 1989

This book is printed on acid-free paper

A CIP record for this book is available from the British Library

Cased edition ISBN 0–12–432562–9
Paperback edition ISBN 0–12–432563–7

Cover illustration

NIH 3T3 cells were incubated in our fixable dye **MitoTracker™ CMXRos** (M-7512), which stains mitochondria red, and then treated with aldehyde fixatives. After the fixed cells were permeabilized with acetone, they were stained with **BODIPY® FL phallacidin** (B-607), which labels actin green, and with **POPO™-1** (P-3580), which labels nuclei blue. This photomicrograph was obtained with a single exposure through the Omega® Optical triple-band filter set (O-5855), available directly from Molecular Probes. Photo contributed by Ian Clements and Sam Wells, Molecular Probes.

Typeset by Fakenham Photosetting Limited, Fakenham, Norfolk
Printed and bound in Great Britain by Mackays of Chatham, PLC, Chatham, Kent

contains large identifiable cells and is consequently, like *Hirudo* and *Aplysia*, a favourite preparation for studying neural mechanisms at the cellular level, and in particular for studying isolated neurons in culture.

helix-coil transition See ***random coil***.

helix-destabilising proteins (single-stranded binding proteins) Proteins involved in DNA replication. They bind cooperatively to single-stranded DNA, preventing the reformation of the duplex and extending the DNA backbone, thus making the exposed bases more accessible for base pairing.

helix-loop-helix A motif associated with ***transcription factors***, allowing them to recognise and bind to specific DNA sequences. Two ***α-helices*** are separated by a loop. Examples: myoblast *MyoD1*, c-*myc*, *Drosophila* genes *daughterless*, *hairy*, *twist*, *scute*, *achaete*, *asense*. Not the same as helix-turn-helix.

helix-turn-helix A motif associated with ***transcription factors***, allowing them to bind to and recognise specific DNA sequences. Two ***amphipathic α-helices*** are separated by a short sequence with a ***β-sheet***. One helix lies across the major groove of the DNA, while the recognition helix enters the major groove and interacts with specific bases. An example in ***Drosophila*** is the ***homeotic*** gene ***fushi tarazu***, that binds to the sequence TCAATTAAATGA. Not the same as helix-loop-helix.

helodermin See ***exendin***.

helospectin See ***exendin***.

helper factor A group of factors apparently produced by helper T-lymphocytes that act specifically or non-specifically to transfer T-cell help to other classes of lymphocytes. The existence of specific T-cell helper factor is uncertain.

helper T-cell See ***T-helper cells***.

helper virus A virus that will allow the replication of a co-infecting defective virus by producing the necessary protein

hema-, hemo- See ***haema, haemo***.

heme See ***haem***.

hemicellulose Class of plant cell-wall polysaccharide that cannot be extracted from the wall by hot water or chelating agents, but can be extracted by aqueous alkali. Includes ***xylan***, ***glucuronoxylan***, ***arabinoxylan***, ***arabinogalactan II***, ***glucomannan***, ***xyloglucan*** and galactomannan. Part of the cell-wall matrix.

hemidesmosome Specialised junction between an epithelial cell and its basal lamina. Although morphologically similar to half a desmosome (into which intermediate cytokeratin filaments are also inserted), different proteins are involved.

hemizygote Nucleus, cell or organism that has only one of a normally ***diploid*** set of genes. In mammals the male is hemizygous for the ***X-chromosome***.

Hensen's node (primitive knot) Thickening of the avian blastoderm at the cephalic end of the primitive streak. Presumptive notochord cells become concentrated in this region. May well be a source of retinoic acid that is acting as a morphogen in the developing embryo.

heparan sulphate (glycosaminoglycan) Constituent of membrane-associated ***proteoglycans***. The heparan sulphate-binding domain of ***NCAM*** is proposed to augment NCAM–NCAM interactions, suggesting that cell–cell bonds mediated by NCAM may involve interactions between multiple ligands. The putative heparin-binding site on NCAM is a 28 amino acid peptide shown to bind both heparin and retinal cells, as well as to inhibit retinal cell adhesion to NCAM. This strengthens the argument that that this site contributes directly to NCAM-mediated cell–cell adhesion.

heparin Sulphated mucopolysaccharide, found in granules of mast cells, that inhibits the action of thrombin on fibrinogen by potentiating anti-thrombins, thereby interfering with the blood-clotting cascade Platelet factor IV will neutralise heparin

zontal transmission, long incubation periods and chronic progressive phases. Visna virus is in this group, and there are similarities between visna, equine infectious anaemia virus and *HIV*.

lentoid Spherical cluster of retinal cells, formed by aggregation *in vitro*, that has a core of lens-like cells inside which accumulate proteins characteristic of normal lens. The cells concerned derive from retinal glial cells.

Lepore haemoglobin Variant haemoglobin in a rare form of ***thalassaemia***; there is a composite δ-β chain as a result of an unequal *crossing-over* event. The composite chain is functional but synthesised at a reduced rate.

leprosy Disease caused by *Mycobacterium leprae*, an obligate intracellular parasite that survives lysosomal enzyme attack by possessing a waxy coat. Leprosy is a chronic disease associated with depressed cellular (but not humoral) immunity; the bacterium requires a temperature lower than 37°C, and thrives particularly in peripheral Schwann cells and macrophages. Only humans and the nine-banded armadillo are susceptible.

leptonema See *leptotene*.

Leptospira Genus of *spirochaete* bacteria that cause a mild chronic infection in rats and many domestic animals. The bacteria are excreted continuously in the urine, and contact with infected urine or water can result in infection of humans via cuts or breaks in the skin. Infection causes leptospirosis or Weil's disease, a type of jaundice, which is an occupational hazard for sewerage and farm workers.

leptospirosis Weil's disease, caused by infection with *Leptospira*.

leptotene Classical term for the first stage of *prophase* I of *meiosis*, during which the chromosomes condense and become visible.

Lesch–Nyhan syndrome A sex-linked recessive inherited disease in humans that results from mutation in the gene for the purine salvage enzyme HGPRT, located on the X-chromosome. Results in severe mental retardation and distressing behavioural abnormalities, such as compulsive self-mutilation.

lethal mutation *Mutation* that eventually results in the death of an organism carrying the mutation.

LETS (large extracellular transformation/trypsin-sensitive protein) Originally described as a cell surface protein that was altered on transformation *in vitro*; now known to be *fibronectin*.

Leu enkephalin A natural peptide neurotransmitter; see *enkephalins*.

leucine (leu; L; 2-amino-4-methylpentanoic acid; 131 D) The most abundant amino acid found in proteins. Confers hydrophobicity and has a structural rather than a chemical role. See Table A2.

leucine aminopeptidase An *exopeptidase* that removes neutral amino acid residues from the *N*-terminus of proteins.

leucine zipper Motif found in certain *DNA-binding proteins*. In a region of approximately 35 amino acids, every seventh is a leucine. This facilitates dimerisation of two such proteins to form a functional *transcription factor*. Examples of proteins containing leucine zippers are products of the *proto-oncogenes myc*, *fos* and *jun*. See also *AP-1*.

leucinopine (dicarboxypropyl leucine) An analogue of *nopaline* found in crown gall tumours (induced by *Agrobacterium tumefasciens*) that do not synthesise octopine or nopaline.

leucocidin Exotoxins from staphylococcal and streptococcal species of bacteria that cause leucocyte killing or lysis.

leucocyte (USA leukocyte) Generic term for a white blood cell. See ***basophil, eosinophil, lymphocyte, monocyte, neutrophil***.

leucocytosis An excess of leucocytes in the circulation.

Z

Z scheme of photosynthesis A schematic representation of the *light reactions* of *photosynthesis*, in which the photosynthetic reaction centres and electron carriers are arranged according to their electrode potential (free energy) in one dimension and their reaction sequence in the second dimension. This gives a Z shape, the two reaction centres (of photosystems I and II) being linked by the photosynthetic electron transport chain.

Z-disc Region of the *sarcomere* into which *thin filaments* are inserted. Location of *α-actinin in the sarcomere.*

Z-DNA See *DNA*.

zeatin A naturally occurring *cytokinin*, originally isolated from maize seeds. Its riboside is also a cytokinin.

zebrafish *Brachydanio rerio*; fish with a transparent embryo making it possible to follow progeny of single cells until quite late stages of development. This, together with the availability of mutant lines, makes it an important preparation for the study of vertebrate *cell lineage*.

zeta potential The electrostatic potential of a molecule or particle, e.g. cell measured at the plane of hydrodynamic slippage outside the surface of the molecule or cell. Usually measured by electrophoretic mobility. Related to the surface potential and a measure of the electrostatic forces of repulsion the particle or molecule is likely to meet when encountering another of the same sign of charge. See *cell electrophoresis*.

Zigmond chamber See *orientation chamber*.

zinc An essential "trace" element being an essential component of the active site of a variety of enzymes. Zn^{2+} has high affinity for the side-chains of cysteine and histidine. Zinc is present in tissues at levels of ca. 0.1 mM, but intracellular levels must be much lower.

zinc finger A motif associated with *DNA-binding proteins*. A loop of 12 amino acids contains either two cysteine and two histidine groups (a "cysteine–histidine" zinc finger), or four cysteines (a "cysteine–cysteine" zinc finger), that directly coordinate a zinc atom. The loops (usually present in multiples) intercalate directly into the DNA helix. Originally identified in the RNA polymerase III transcription factor TFIIIA.

zipper See *leucine zipper*.

zippering Process suggested to occur in phagocytosis in which the membrane of the phagocyte covers the particle by a progressive adhesive interaction. The evidence for such a mechanism comes from experiments in which capped B-cells are only partially internalised, whereas those with a uniform opsonising coat of anti-IgG are fully engulfed.

ZO-1 High molecular weight protein (225 kD in mouse, 210 kD in MDCK cells) associated with *zonula occludens* (tight junction) in many vertebrate epithelia. *Cingulin*, which is distinct, is found in the same region.

zona pellucida A translucent noncellular layer surrounding the ovum of many mammals.

zone of polarising activity The small group of mesenchyme cells in avian limb buds that is located at the posterior margin of the developing bud and that produces a substance, possibly retinoic acid, which provides positional information to the developing limb bud.

zonula adhaerens Specialised intercellular junction in which the membranes are separated by 15–25 nm, and into which are inserted microfilaments. Similar in structure to two apposed *focal*